FACULTY OF ARTS

CARNAP AND THE UNITY OF SCIENCE
The intellectual and moral formation of a science-technology generalist: a case study

SAMULI SALMI

Division of Theoretical Philosophy
Department of Philosophy, History, Culture and Art Studies
Faculty of Arts
University of Helsinki
Helsinki, Finland

Academic dissertation

To be presented, with the permission of the Faculty of Arts of the University of Helsinki, for public criticism in auditorium XII, Fabianinkatu 33, on September 15th, 2012, at 12 o'clock noon.

Helsinki 2012

Author's contact information: email: samuli.salmi@helsinki.fi

Supervisor: Professor Matti Sintonen, Ph.D.
Department of Philosophy, History, Culture and Art Studies
University of Helsinki

Reviewers: Professor Jaakko Hintikka, Ph.D.
Department of Philosophy
University of Boston

Professor Thomas Uebel, Ph.D.
School of Social Sciences
University of Manchester

Opponent: Professor Friedrich Stadler, Ph.D.
Faculty of Historical-Cultural Sciences and
Faculty of Philosophy and Educational Sciences
University of Vienna

ISBN 978-952-10-8245-0 (printed version)
Helsinki 2012
Unigrafia Oy

ISBN 978-952-10-8246-7 (pdf version)
http://ethesis.helsinki.fi
Helsinki 2012
Helsingin yliopiston verkkojulkaisut (Unigrafia Oy)

Acknowledgements

First of all I am indebted to Professor Matti Sintonen whose politeness and support I have found invaluable in the midst of the quite solitary undertaking of mine. He has encouraged (tolerated?) me in my pursuit of a number of different lines of enquiry — often seemingly disparate ones — that I have suggested as being relevant for one another. This attitude has resonated with my inclination towards a pluralistic approach in philosophy. However, all being said and done, the thing that I value most is his friendliness. Professor Ilkka Niiniluoto has also supported my work on critical occasions. For this I am most grateful.

The *animateurs des idèes* at *Institut Wiener Kreis*, Professor Friedrich Stadler, Dr. Robert Kaller, and Dr. Karoly Kokai kindly permitted me to study the materials in the Vienna Circle Archive. They have been of enormous help. I am also indebted to the Program manager of the science archives, Godelive Bolten, at *Noord-Hollands Archief* for sending me the material appertaining to the Neurath-Carnap Correspondence.

I hereby acknowledge the financial support of *The Research Foundation of the University of Helsinki*, *Koneen säätiö*, and *The Department of Research Affairs at the University of Helsinki*.

On a more personal note, I am indebted to my parents, Kauko and Anita Salmi, and my sister Sanna Mattila, who have always exemplified unwavering trust towards me and my aspirations. Mikko Mattila has always been ready to participate in the many supportive tasks relating to the usual life of a family with little kids. Thank you for your generosity. Timo, Anne & Tuula Karlsson have shared an interest in my somewhat hermetic undertakings. It has been a pleasure to delight in the beautiful scenery and serenity of Trullevi, thanks to your cooperation and warm hospitality. Mr. Arto Koskinen has contributed towards the completion of this project more than he can be aware of. His exemplary intellectual and moral integrity have encouraged me to pursue my interests despite all the contingencies that occasionally befall a citizen of this planet. For his support I am extremely grateful. Mr. Matti Helaste has been a solid rock support in spite of the lamentable distance between our present home towns. People that possess similar gifts for providing assurance and encouragement are rare. Thank you. I express my warmest thanks to Mr. Veli-Pekka Suomalainen for interesting conversations ranging over a wide spectrum spanning astronomy, physics, mathematics, ophthalmology, music and literature, for charitably putting at our disposal the Hammond A-100, and most of all for his friendship. Anna-Stiina Tarkka, Jani Indrén, Jouni Jarnamo & Mika Kaalikoski, as well as Ilkka Joronen, Okko Kivikataja, Olli-Pekka Kankimäki & Mikael Seire, thank you for a constant stream of rhythmic memories, I am indebted to you all. However, I doubt that I can offer you this book to beguile the tedious hours of the night . . . Thank you for sharing with me the wonders of music! The same goes without saying to the members of my Hong Kong -based combo, "The Band with No Name": Dr. John Chung, Mr. Francis Wong, Mr. Desmond Yau, Mr. Samson Cheung and Mr. Daniel Cabane. I will always cherish the memory of our sessions at Tsim Sha Tsui and Fo Tan! The list would be incomplete without Dr. Daniel Adriano-Silva, a biochemist, cosmopolitan, gentleman, philanthropist and great friend. Without you the world would be a much more miserable place.

The final words of gratitude I express to my family. My wife, Mira, has made the whole process tolerable with her love and support. Her admirable flexibility has made working on this project on odd hours possible and the occasional retreats to the summer house a most welcome luxury. I dedicate this book to my wonderful children, Leo & Linnea, who are not quite yet in a position to grasp why their father has been buried in his library like a mouse in a cheese.

Trullevi, Kokkola June 4, 2012
Töölö, Helsinki June 29, 2012

Samuli Salmi

Πάντες ἄνθρωποι τοῦ εἰδέναι ὀρέγονται φύσει. [1]

[1] "All humans desire to know by nature." (Aristotle, *Metaphysics* I.980a21)

Samuli Juhani Salmi
University of Helsinki, 2012
Abstract

This dissertation concentrates on a particular exemplification of the ideal of the unity of science in the history of twentieth-century philosophy. Taking Rudolf Carnap (1891–1970) as an exemplar of a scholar whose work in philosophy of science was at bottom motivated by the ideal of a unified conception of science, it attempts to distillate the essential characteristics and methodological significance of such a conception by a combination of historical and systematic analysis. Given the conspicuously holoscopic character of Carnap's philosophical orientation, there arises an interesting question about the relation of his work to that of other prominent "seekers of the wider view" in the history of philosophy (and history of science). On a more general level, we ask what kind of intellectual and moral characteristics are associated with a scholar who is motivated by the unification of science. Making it explicit: if a coherent representation of a unified conception of science is conceivable, what kind of normative criteria can then be applied to a scholar and his actions? In other words, what are the external and internal qualifications of scholar's vocation under the unified conception of science? On our view, the question can be answered satisfactorily only when one supplements the philosophical approach of Carnap with a comprehensive view on man, the formation of a (scientific) self and the relationship between the concepts of the moral and the scientific.

Keywords: Rudolf Carnap, philosophy of science, unity of science, universalism, ethics, self

Contents

INTRODUCTION

> Halten wir uns bereit, die Gewöhnungen unseres Tages abzustreifen und das Vergangene wieder als Hort verschertzer und vergessener, aber lebendigen Möglichkeiten zu ehren! Nur so entrinnen wir nämlich der würdelosen Despotie des Zeitgeists, nur so gewinnen wir jene Freiheit, die einzig der Raum der Gesichte gewährt. Wir haben uns von Nachteil der Historie überzeugen lassen und viel zu spät bemerkt, wie eng sich unser Horizont einschränkt, wie kümmerlich unser Wachstum ist, sobald wir in der Geschichte nichts als Widerspruch oder Bestätigung suchen, statt, vorerst von nichts als Neugier geleitet, unbekümmert und selbstvergessen, in ihre Gefilde auszuschweifen.
>
> – Emil Staiger, *Goethe*, vol. I

Rudolf Carnap (1891–1970) is generally acknowledged as one of the most important philosophers of the twentieth century. In the last twenty years or so, the philosophical community has witnessed a veritable renaissance of Carnap scholarship.[2] The fruits of this collective effort have now been available as high-quality monographs and articles that provide a comprehensive and balanced picture of the emergence, development and reception of various aspects of Carnap's philosophy. This body of work is both informative and enlightening, ruminating in areas that lie outside the ring of special questions that were studied by the circle of specialists that followed in Carnap's footsteps around the middle of the twentieth century. From the 1950s on, Carnap's overt interests were mainly focused on inductive logic. His pupils, advocates and acolytes were inspired by

[2] The range of scholars that have contributed to this new wave of historically and scientifically informed study of Carnap, the Vienna Circle and logical empiricism is extensive and exemplifies the truly cosmopolitan spirit of enquiry within this particular domain. The scholars and works that have most of all opened up new vistas and lines of interpretation in this vein are Steve Awodey [Awodey & Carus (2001), (2003), (2004), (2007a), (2007b)], [Awodey & Klein (2004)], [Awodey & Reck (2002a), (2002b)], André Carus [Carus (1999), (2001), (2004), (2007)], Alberto Coffa [Coffa (1987), (1991)], Richard Creath [Creath (1982), (1990), (1996)], Hans-Joachim Dahms [Dahms (1994), (2004)], Michael Friedman [Friedman (1999)], [Friedman & Creath (2007)], Manfred Geier [Geier (1992)], Guillermo E. Rosado Haddock [Haddock (2008)], Juha Manninen [Manninen (1995), (2001), (2002a), (2002b), (2002c), (2002d), (2002e), (2003), (2004a), (2004b), (2009a), (2009b)], Thomas Mormann [Mormann (2000), (2003), (2007)], Erich H. Reck [Reck (2002), (2004), (2007)], George A. Reisch [Reisch (1991), (2005)], Alan Richardson [Richardson (1998), (2007)], [Richardson & Uebel (2007)], Thomas Ricketts [Ricketts (1996)], Thomas Ryckman [Ryckman (2005)], Friedrich Stadler [Stadler (1997), (2007)], Klemens Szaniawski [Szaniawski (1989)] and Thomas Uebel [Uebel (1991)]. Naturally, this list is incomplete. However, it forms a kind of a minimal set by the help of which one might start to find one's bearings in the domain of Carnap studies.

Carnap's work in this area and immersed themselves in the numerous technical questions springing forth from their study. Much of this work related to the problem of formulating a mathematical framework for rational decision making, conceived in the twofold sense of comprising both statistical inference as used in the evaluation and estimation of hypotheses, and decision theory as a framework for guiding our actions in the face of uncertainty. Carnap had joined the philosophy department at UCLA in 1954, where Hans Reichenbach (1891–1953) had held the chair in philosophy before him. There Carnap continued to work on the foundations of probability, thermodynamics and inductive logic, collaborating with many younger colleagues, until his death in 1970. The results of this work are compiled in the impressive two-volume *Studies in Inductive Logic and Probability* [Carnap & Jeffrey (1971); Jeffrey (1980)] Antecedent to this intense period of work in inductive logic and foundations of probability Carnap had immersed himself in the study of semantics. This work, culminating in the book *Meaning and Necessity* [Carnap (1947)] was characterized with a technical sophistication that had, by then, come to epitomize Carnap's highly distinctive style in philosophy. In the University of Chicago where Carnap was a professor of philosophy from 1936 to 1952, he was regarded by the faculty as a kind of a technician, not engaging in philosophy at all.[3] Intense concentration on overtly technical questions during the latter part of his career, eschewing any attempt to disclose the underlying motivation or broader ideological background of this technical work, contributed towards the distorted view of his philosophy in particular and of logical empiricism in general that emerged in the general consciousness during the 1960s. The radical shift in intellectual atmosphere during that time, fueled in part by the geopolitical tensions as well as social and political surges, and most prominently by the radical attitudes of the idealistic young of the 1968, made it evident that the seemingly sterile doctrines of scientific philosophy that had been conceived in Vienna in the 1920s and 1930s were of no interest to the youth that attempted to find their place in the chaotic and tumultuous world. Influenced by the historicist turn in the philosophy of science, that was mainly catalyzed by Thomas Kuhn's book *The Structure of Scientific Revolutions*, the Academe, so the story goes, was liberated from a positivist or objectivist conception of science that privileged the so called 'hard' sciences at the expense of humanistic or social disciplines. All of these factors together contributed to the fact that from the 1960s and 1970s on some monographs and articles referring to the history of logical empiricism and its philosophical import perpetuated a curious view of its basic doctrines and especially of Carnap. In these accounts (inspired, as pointed above, mainly by Kuhnian ideas in history and philosophy of science) the intellectual portrait of Carnap was painted with cubistic strokes resulting in a picture of a stern technocrat interested only in a very narrow field of problems. (Although it has to be said that as far as the manifest image of Carnap is concerned, this assessment was justified.) Carnap's famous dictum in the *Logische Syntax der Sprache* [Carnap (1934a)] that "philosophy is the logical syntax of the

[3] His sometime pupil Richard Jeffrey has related that: "What Carnap was doing wasn't seen as philosophy by most of the faculty; they regarded it as a kind of engineering. And Carnap seemed perfectly content with that description. [...]" [Carus (2007), 36n] Indeed, within the faculty only Charles W. Morris (1901–1979) shared with him the conviction about the primacy of logic and science as the cornerstones of rigorous philosophy. Their colleagues included Richard McKeon (1900–1985), Mortimer Adler (1902–2001), Charles Hartshorne (1897–2000), and Manley Thompson (1917–1994), all of whom were more or less inspired by the then current variety of American pragmatism.

language of science ″ accompanied with the earlier vehement manifesto of *Scheinprobleme in Philosophie* [Carnap (1928b)] that culminated in the *veto* to eliminate philosophical problems that had traditionally been conceived as having perennial value,[4] made Carnap appear as a scientific purist on par with such hard-boiled positivists as Ernst Haeckel (1834–1919), Dimitri Pisarev (1840–1868) or Wilhelm Ostwald (1853–1932). The technical machinery of his publications, which to the uninitiated must have appeared as a spirit-sapping formalism which only the few and devoted had the patience or perseverance to master, surely helped to conceal his status as a philosopher of first rank in the eyes of the wider public interested in philosophical questions. But the impression of Carnap as a technocrat obscured the underlying motivation driving Carnap in his particular 'language engineering' projects. Indeed, the projects that Carnap was engaged with during the last twenty years of his life were very much in the service of the radical ideas that he had conceived already in 1931–1932. But unfortunately for the philosophical community and culture at large, because of the adverse circumstances, this failed to come across in the America, where the intellectual and cultural atmosphere was very much different from the exuberant and lively *milieu* of Vienna of the 1920s, where logical empiricism was born. Still, Carnap himself has to be taken as accountable for not having made explicit the connections between his philosophical program and his political activity which he continued even while at the United States. In keeping with his earlier activities in Europe, he was involved in radical politics, and was, mainly for this reason, extensively spied on by FBI.[5] The conditions of McCarthyism in the 1950s made Carnap careful to not associate his philosophical work with anything that might compromise his freedom in the way of attracting unnecessary attention from state authorities or university administration. All in all, given the various social and political tensions characterizing the American society in the middle of twentieth century, it is not surprising that Carnap suspended from any public expression of the ambitious and radical utopianism that underlay the philosophy of the Vienna Circle.

[4] I have in mind, naturally, problems that are usually labeled *metaphysical*. Not that Carnap alone was influential in propounding such an antimetaphysical view; as a matter of fact, some of his colleagues in the Vienna Circle showed a much more militant attitude against such questions (Otto Neurath is an example *par excellence*). Carnap was antimetaphysical only in the sense that he denied the possibility of meaningfully engaging in metaphysical discussions using the medium of scientific (philosophical) language. Emotionally he was most sympathetic to expressions of the *Lebensgefühle* in the form of poetry and music, for example. Indeed, the most striking simile Carnap ever produced concerning this question, is the following given in [Carnap 1931 [Carnap 1959, 80]]: "The harmonious feeling or attitude that the metaphysician tries to express in a monistic system is given clearer expression in the music of Mozart. And when a metaphysician gives verbal expression to his dualistic-heroic attitude towards life in a dualistic system, is it not perhaps because he lacks the ability of a Beethoven to express this attitude in an adequate medium? *Metaphysicians are musicians without musical ability*." Just to see how much the general atmosphere surrounding these questions has relaxed even within the confines of the analytic 'school', it is illuminating to relate how Michael Dummett empathizes with the layman's expectations that philosophers should answer "deep questions of great import for an understanding of the world". According to Dummett, an inventory of such questions would include at least the following: "Do we have free will?", "Can the soul, or the mind, exist apart from the body?", "How can we tell what is right and what is wrong?", "*Is* there any right and wrong, or do we just make it up?", "Could we know the future and affect the past?", "Is there a God?". Although these concerns are "disconcertingly remote from the writings of the analytical school", as Dummett emphasizes, "if philosophy does not aim at answering such questions, it is worth nothing". [Dummett 1991, 1]

[5] George A. Reisch has established these facts in his recent book *How the Cold War Transformed Philosophy of Science* [Reisch (2005)].

But who, then, was Carnap? An answer to this question has begun to take definite shape through the collective efforts of many scholars (some of whom were mentioned above). Rudolf Carnap, the philosophical leader *de facto* of the Vienna Circle is no longer considered as a solitary character of passing interest that he was made into in the 1960s when the antagonism towards logical empiricism and everything that it was taken to represent reached its high-point. Indeed, the contrast with the picture of the early reception history of Carnap is conspicuous, to say the least. This is due, in part, to the acknowledgement by the recent scholarship of the influence of a number of seemingly incompatible currents of German and Austrian thought on Carnap's philosophy. Most prominent among these influences are the neo-Kantian movement, particularly the Marburg school represented by Hermann Cohen (1842–1918), Paul Natorp (1854–1924) and Ernst Cassirer (1874–1945), the pragmatically oriented fictionalism of the neo-Kantian Hans Vaihinger (1852–1933), the phenomenalism of Ernst Mach (1838–1916), the *Lebensphilosophie* and philosophical pedagogy of Herman Nohl (1879–1960), the phenomenology of Edmund Husserl (1859–1938) and the logic and philosophy of Gottlob Frege (1848–1925). Although the great importance of Frege's thought for Carnap's philosophy has been emphasized by a range of scholars from the 1960s on, and even by Carnap himself, the impact of Edmund Husserl has demonstrably been of equal rank, at least upon Carnap's early philosophy.[6] Disclosing the broader German background of Carnap's thought in its totality is a welcome palliative against the one-sided emphasis on the anglo-saxon analytic tradition, foremost Frege, Wittgenstein and Russell, as the background of Carnap's philosophy. It was precisely this one-sidedness that led to the caricatures of both Carnap and logical empiricism that were so influential in the post-war academic philosophy. That Carnap was a first-rate philosopher of science, no one seems to dispute, but that he also perceived wider intellectual horizons than what is made explicit in his technical work, escapes many.

Carnap's philosophy, fueled as it was with an intense motivation to analyse the most essential conceptual tools used in science, to clarify the procedures of introducing hypotheses and the methods of confirmation within different compartments of science and to make explicit the complex edifice constituted by the knowledge provided by different disciplines, building upon the foundations of the exact specification of the different languages used in them, including the menagerie of logics, the vast domain of mathematics and the different interpretations of probability, was also nourished by a more humane source of questions, worries and desires. A central problem which underpins Carnap's philosophical program is the relevance of philosophical and scientific thought to the practical worries of mankind, in effect, to *life*. Despite Carnap's apparent focus on research topics that directly bear on the most theoretical and abstract work in the sciences, he was, at least implicitly, striving for an integrative conception of the significance of science and systematic knowledge for the everyday life of society and the individual. He was essentially attempting to work out the implications of the programmatic manifesto of the Vienna Circle, in the conception of which he played a part, published under the name *Wissenschaftliche Weltauffassung: Der Wiener Kreis* ("Scientific World Conception: The Vi-

[6] This is a rather recent finding which I shall have occasion to comment upon in this work. The most prominent scholars who have studied Carnap's connection with Husserl are Guillermo E. Rosado Haddock [Haddock (2008)], Thomas Ryckman [Ryckman (2007)], Jean-Michel Roy [Roy (2004)] and Sahotra Sarkar [Sarkar (2003)].

enna Circle") in 1929. In place of the utopian verse of that manifesto he put concrete proposals for improving the conceptual practices of science (the most important part of this work concerning the notion of confirmation in the empirical sciences and different systems of inductive logic), with a view that such proposals would ultimately contribute towards a wider appreciation of science and the scientific world view, by removing the opaqueness of scientific language and unifying different domains of scientific work by eliminating unintended ambiguity and idiosyncratic expressions from the vocabulary of the special sciences. These conceptual clarifications, or 'conceptual engineering tasks', formed only a part of the overall mission that lay implicit in the particular projects that Carnap delved in. Carnap was very much inspired by ideas that began to emerge in the *fin-de-siècle* central Europe, mainly in Berlin and Vienna, ideas fermented in the atmosphere of modernism that spanned the diverse fields of philosophy, literature, poetry, visual art, architecture, music, religion and politics. The variety of cultural, social and political influences in "Wittgenstein's Vienna" constituted a unique blend, resulting in an exuberant blossoming of a wide range of cultural activity, novel ideas and forms of expression, not to mention political scuffles.[7] The Vienna of the 1920s where Carnap spent the most fruitful and creative years of his philosophical activity was thus permeated with a myriad of influences that constituted the essence of modern Europe. The ideas that foremost inspired the members of the Vienna Circle, and Carnap in particular, had their roots in the Enlightenment. Quite similarly, in the spirit of the French *Encyclopédistes* (foremost d'Alembert, Condorcet and Diderot) the main goal of whom was to make scientific knowledge available to everyone in the true spirit of intellectual equality, pursuing an ideal of a cosmopolitan republic of letters, the constitution of an architectonic for a unified science, with mutual respect and openness between colleagues in an international scale, was one of Carnap's ideals. The explication of concepts, the attempt to render them more conducive to scientific work and to facilitate the progress of science in general while aspiring simultaneously to link the diverse findings and results of scientific investigation more appropriately with practical life became the *modus operandi* of his mature philosophy. Similar questions formed the focus of many other European scientists in the earlier part of twentieth century. Wolfgang Köhler had boldly addressed the question about "the place of value in the world of facts" in the 1930s when general opinion of science was quite reserved, even hostile. The post-World-War societies, on the contrary, were infused with exuberant enthusiasm and trust in the utility of science, especially in America (where the growth of technology it spurred in an unforeseen manner had played a prominent role in winning the war) and in Germany (where the experience of the catastrophe of the Nazi regime with its anti-scientific romanticism made post-war governments generally very favorable towards enlightened and scientific views). In spite of these developments there gradually arose opinions underlining the truculent implications of 'scienticism' among public. The Cold War (the first truly global civil war) and the tragedy of Vietnam, as well as ecological catastrophes like the Seveso disaster in 1976, at the latest, opened the eyes of the new generation to evaluate critically the conceptual and moral foundations of the post-war way of life which was largely based on a scien-

[7] The atmosphere of Vienna in the first third of twentieth century with its diametrically opposite ideologies and the tangible tension between the attitudes of the *Alt Wien* and the *Jung Wien* has been masterfully depicted by Allan Janik and Stephen Toulmin. [Janik & Toulmin (1973)]

tific world conception. The impending global ecological crisis could, according to one view, be seen as a consequence of the gradual opening of a cleft between physical being and consciousness, the historical consciousness of the distinction between a pure consciousness and nature without subjectivity, to be found in its most extreme form in the Cartesian dualism between *res cogitans* and *res extensa*.

How was the scientifically oriented world vision of Carnap's related to these socio-psychological changes? In the eyes of the new generation, it no doubt appeared in a suspicious light, representing a narrow-minded and obsolete view on the nature of human cognitive projects in general, and philosophy in particular. This impression rested on a one-sided view of Carnap's philosophy, however. Carnap's overall conception of philosophy as one among the cognitive disciplines was very similar to the one of Köhler. Like Köhler, Carnap was brought up in an environment where science and all the other branches of human knowledge were held in high esteem. In Germany, as a young student, he came under the influence of many eminent scientists, philosophers and artists, whose work comprised in his eyes achievements of highest value. In Germany the governments considered it as one of their noblest and honorable duties to support all forms of research and fine arts that were seen to contribute towards the overall spiritual well-being of mankind. Coming from such an environment, Carnap very clearly perceived the intricate, but enormously important, task of reconciling the content of highest scientific and artistic achievements with the practical problems of daily life. Carnap's motives were thus political in the widest sense; he wished to find ways to transform society and its established policies that he conceived as opposite to the ideal that appertained to the enlightened and tolerant thinking associated with the scientific world conception. This was a markedly pragmatic ingredient in Carnap's philosophical temperament. It is a mistake to see in it a sign of slender utopianism; it really was an exemplification of his deep feeling of responsibility that he thought every scientist or scientifically oriented human being was endowed with. In perpetuating his ideas, Carnap was not content with mere philosophical elucidations [*philosophische Erörterungen*]. The programmatic *ethos* of his philosophy was inextricably entwined with a concrete and pervasive knowledge of the details of the problems he tackled. Indeed, he might well have applauded the words of William James, who took it as a necessary desideratum in all his work to "forge every sentence in the teeth of irreducible and stubborn facts." Carnap thought that the reconciliation he sought for could be accomplished by a completely new philosophy. This philosophy he conceived as a science, or at least, as a 'scientific philosophy', inspired by the systematic work of the great scientists of the twentieth century, including Einstein, Poincaré, Weyl, Pauli, Schrödinger, Heisenberg, and others, whose exemplary work could be used as a model in devising the foundations of this new philosophy.

When Carnap was interviewed in 1964 by Dr. Hochkeppel, he was asked what is meant by the expression "scientific philosophy" of which he was generally considered to be one of the main representatives. Carnap came up with a thoughtful and qualified answer, as always:

> Ich bin nicht ganz sicher ob das die beste und glücklichste Bezeichnung ist, aber wir haben sie oft verwendet, und ich glaube, sie hat doch eine gewisse

> Berichtigung. Natürlich, man muß sie nicht mißverstehen, als wäre die Philosophie wissenschaftlich in dem Sinne, daß sie genau dieselbe Methode und denselben Inhalt hat, wie die Wissenschaft. Das ist sicherlich nicht der Fall. Da ist ja ein deutlicher Unterschied. Die Wissenschaft hat die Aufgabe, die Fakten in der Natur aufzusuchen, zusammenzustellen, zu vergleichen, zu erklären und so weiter und uns dadurch ein Bild der Natur zu geben. Die Philosophie dagegen soll sich in das Gebiet der Wissenschaft nicht einmischen. Manchmal haben das Philosophen getan. Das halte ich für unberechtigt. Sie erinnern sich, solche kosmologischen Philosophen taten das. Die ganze Kosmologie ist aber Sache der Wissenschaft, das ist nicht unsere Sache.
> Trotzdem, glaube ich, kann man sagen, daß die Philosophie wissenschaftlich ist oder sein sollte, und wir bemühen uns, sie dahin zu bringen; aber nur in dem Sinne, daß sie dieselben Forderungen stellt, nämlich Standards von Objektivität und Rationalität in der Argumentation. Wir glauben, daß in der traditionellen Philosophie, besonders in der Metaphysik, oft in einer, so könnte man beinache sagen, unverantwortlichen Weise, jedenfalls un einer sehr subjektiven Weise, versucht worden ist, etwas darzustellen, das dann als Erkenntnis gelten sollte. [Carnap (1993), 133–134]

The members of Vienna Circle — at least the more radical and enlightened of them — were really attempting to redefine the criteria of objectivity and rationality — in their cognitive, emotional and conative aspects — that underpin the aspirations of mankind. And this is what Carnap's work in philosophy really amounts to: it is an attempt to figure out the broad outlines of human rationality, providing a radically new alternative to the received view deriving from Kant and his predecessors.

The evolving conception of analysis in Carnap's philosophy is the thread that connects all the aspects of his work together. Especially pertinent to this conception is Carnap's demarcation between logical and non-logical expressions in any given system of logic, and the criterion of *logicality* in general which is exemplified in this demarcation. The centrality of logic and mathematics in Carnap's philosophical program is seen to be based on the tenet that they constitute the largest compartment of science where one can define validity by syntactic means alone. Although Carnap seems to abandon this view in the early 1940s, adopting a semantic approach to the question of logicality, the original idea remained an essential element in Carnap's thought. Indeed, the need for a definition of logical expressions becomes all the more urgent as one makes a transition from strict logical universalism to logical pluralism, as Carnap in fact did. Thus, on the one hand, along with the different senses of analysis there emerge different senses of logicality and invariance the most explicit rendering of which Carnap provides in the *Logische Syntax der Sprache*. On the other hand, from the *Logische Syntax der Sprache* on, the Principle of Tolerance constitutes the *fundamentum philosophiæ* in Carnap's thought. The liberation that is marked by the transition from the universal logical framework (epitomized by the ramified type-theory of Russell and Whitehead) to the pluralism of many different logics ("the open sea of free possibilities") constitutes one of the watersheds of Carnap's intellectual development. These two issues are naturally inextricably entwined. Towards the end of 1940s the Carnap's task becomes manifestly pragmatic, i.e., one of making available different tools of communication and expression, tools that could be utilized in every-day scientific work. The broader implications of the Principle of Tolerance, however, are not overtly discussed. Carnap steadfastly continues to work on the specific problems that he finds the most relevant for the explicit expression of the scientific philosophy. The method of explication which becomes the hallmark of Carnap's philosophical work in the late 1940s, however, is clearly undergirded by the Principle. Carnap illustrated the use of the method and discussed its underlying motivation in the first chapter of *Logical Foundations of Probability*. But he never discussed explicitly the broader questions and problems that the method would enable one to formulate accurately. Indeed, the general questions about the practical utility of theoretical knowledge and the evaluation of different proposals for a theoretical description of a practical problem within a social or political framework, which could be fruitfully addressed with his method of explication, were never elaborated in his published writings. That Carnap's thought ramified into areas outside the immediate concerns of philosophy of science to encompass such broader questions is hinted at by the striking parable of Herbert G Bohnert:

> For a philosopher to concentrate very hard on any one thing, even if that thing be generality, is, for some, a sign of narrowness, the mark of the specialist. A.J. Ayer once divided all philosophers into pontiffs and journeymen, with Carnap as the chief example of the latter. And Richard McKeon sorted them into holoscopic and meroscopic types, with Carnap as the chief example of the part-peerer. Ayer's division was more kindly to Carnap than McKeon's but both missed Carnap's scope. [...] Perhaps the only way I can convey my counterimpression is by a parable of my own.
>
> Picture a *very* holoscopic mind. Suppose it is a very powerful one. After a survey of the whole scene it would, of course, form plans. The plans would require deeper study in certain areas. This deeper study would reveal broader promises and puzzles. Interrelationships would be perceived. Plans and studies, by interaction, would quickly become global. Science would have to be unified, language systematized, the foundations of reasoning and experience scrutinized. Many specialized, meroscopic jobs would have to be done. Some could be done best by the mind itself — like constructing the needed overall conceptual framework — but time is limited. Minds must organize. A journal must be started. A manifesto of the new plan must be issued, congresses scheduled, an encyclopedia planned. Delays must be expected, of course. Wars. The interaction of minds is uncertain; the interaction of groups is unguided. In the meantime, the mind applies itself to those very special but very basic jobs which few but it can perceive as essential and promising. This brings the mind fame as a great specialist.
>
> At some point, of course, the whole plan is seen as unlikely to progress beyond its most initial phases in the time the mind sees as available to it. But then it was never unaware of probabilities. As a holoscopic mind, it understands its predicament perfectly. It still likes the plan. It proceeds in its

painstaking work as if it had millennia. [Bohnert (1975) [Hintikka (1975b), XLIII–XLIV]]

Despite the numerous *post mortem* reviews explaining the reasons behind its ultimate failure, Carnap's program of explication, it seems to me, has very much pertinence and philosophical staying power even today. These reviews have for the most part taken the singular — and obvious — shortcomings of the specific suggestions Carnap made as evidence for arguing for the overall impossibility of carrying through a project that Carnap envisaged. But surely the specific failure of one inadequate construction is not tantamount to a refutation of the possibility of devising a better and more coherent construction.

This dissertation concentrates on a particular exemplification of the ideal of the unity of science in the history of twentieth-century philosophy. Taking Rudolf Carnap (1891–1970) as an exemplar of a scholar whose work in philosophy of science was at bottom motivated by the ideal of a unified conception of science, it attempts to distillate the essential characteristics and methodological significance of such a conception by a combination of historical and systematic analysis. Given the conspicuously *holoscopic* character of Carnap's philosophical orientation, there arises an interesting question about the relation of his work to that of other prominent "seekers of the wider view" in the history of philosophy (and history of science). On a more general level, we ask what kind of intellectual and moral characteristics are associated with a scholar who is motivated by the unification of science. Making it explicit: if a coherent representation of a unified conception of science is conceivable, what kind of normative criteria can then be applied to a scholar and his actions? In other words, what are the external and internal qualifications of scholar's vocation under the unified conception of science?

In the first part of the dissertation we provide a general account of the problem's background in the intersection of intellectual history and systematics. In the first chapter main emphasis will be put to the dialectic between agent-based and structural explanations in historiography. The survey of a few exemplars of models of historical explanation is intended to provide a background framework for discussing the relation between descriptive analysis and analysis of values. In as much as our modern scientific world conception and the general, essentially human, consciousness of the domain of validity seem to be in a fundamental conflict, a philosophical clarification of the issues that depend on this fundamental distinction is contingent on having proper tools at its disposal. Indeed, it is necessary to acknowledge – with respect to *both* scientific knowledge and moral positions – that the issues of genesis and validity have little in common. Both the image of nature, built upon the masses of scientific and technological knowledge gathered, and the modern conceptions of the moral have developed in the course of history. The lesson that historicism can teach us is the possibility to adopt a symmetrical attitude with respect to the status of the questions of genesis and validity within these (very different) domains. This symmetric attitude enables us to see that the validity of a theory or position (in science or in moral philosophy) cannot depend on the diachronic aspects of its genesis. Rather, it is precisely the case that the late appearance of certain scientific theories and certain moral positions is an index that they are complex and presuppose a great deal genetically, and this is seen to be a common feature of all good theories. Thus, if we would like to be able to approach the evolution of these ideas from a general perspective, we have to acknowledge their fundamental ontological difference and adopt a variety of tools to study these domains. I present four different approaches to the study of historical phenomena that appertain to the themes of this dissertation.

In the second chapter we provide a synopsis of the important thematic about the relationship between morals and science. After a brief examination of the concepts of the moral and the scientific, we proceed to give an account of the concept of *scientific self* which acts as a kind of normative meta-concept co-ordinating the interaction between the *epistemic* and the *ethical* requirements appertaining to the education and professional formation of a scientist. From a historical perspective it is easy to to see that the intension of the concept of scientific self varies according to the contingent factors such as the external conditions of education and the requirements set by new experimental techniques, but the essential, axiologically relevant, internal determinates of the concept are seen to accumulate over time in a conservative manner. Especially interesting here are the determinates that can be traced back to the complementary intellectual traditions of Enlightenment and Romanticism. One of the most important exemplifications of an articulated conception of scientific self can be found in J. G. Fichte's "*Vorlesungen über die Bestimmung des Gelehrten*" of 1794. In these lectures Fichte develops a beautiful – and still highly relevant – conception of the true goals of a scholar as well as the qualifications he must fulfill to attain those goals. From Fichte we turn to study the history of one particular intellectual virtue that has direct relevance for the questions tackled in the second part of the dissertation, viz. tolerance.

In the third chapter we focus on the importance of *a priori* knowledge for both ethics and science. These themes are developed only in their barest outlines in order to provide some theoretical support to the fundamental philosophical thesis of the dissertation concerning the distinction between *Is* and *Ought* and its relevance for the question of the unity of science. We will briefly touch upon the question about the relationship of *a priori* and empirical knowledge in ethics, and provide a brief synopsis of the relevance of the distinction analytic/synthetic in this domain. With respect to the discussion about the dynamics between the *a priori* and empirical elements we provide criteria for decisions that are morally right. Finally we address the ontologically crucial problem about the moral element in man and present – with a view to the Enlightenment virtues – a synopsis of the process of the dissociation of the concept of the moral from the concept of the scientific. We describe the characteristics of ethical impulse in modern times and the quite idiosyncratic view on morals and especially on moral justification advocated by the members of Vienna Circle. We will see how the dissociation of the moral from the domain of the rational discourse inevitably results in the philosophically poverished stance of moral non-cognitivism which Carnap maintained throughout his career.

In the second part of the dissertation we can finally address the adduced problem in its particular ramifications in the philosophy of Rudolf Carnap. Given this general problematic, we attempt to vindicate the underlying overall motivation of Carnap's philosophy and to reconstruct the architectonic of Carnap's systematic thought in the light of the most recent research. One of the main tasks is to evaluate the coherence of interpretations provided in the research literature which place Carnap in the continuum of thinkers that are, in some sense, committed to the ideals and values of Enlightenment. The most explicit rendering of this line of thought is the recent monograph by A.W. Carus [Carus 2007] which puts Carnap's method of explication on center stage. I critically examine this line of interpretation indicated by Carus and explore more deeply its historical dimensions. Over and above the interpretation of Carus, we assess to what extent Carnap's philosophical program fulfills the criteria that are imposed upon it by the requirement of an Enlightenment conception of unified science.

The central significance of logic and mathematics in Carnap's philosophical program is seen to derive from the fundamental conception of Carnap that within the total system of knowledge logic and mathematics are performing the essential role of supplying the forms of concepts, statements, and inferences, forms which are then applicable everywhere, hence also to non-logical knowledge. Therefore, the demarcation between logical and non-logical expressions, along with the Principle of Tolerance and logical pluralism, constitutes one of the central strands of Carnap's thought. Indeed, the Principle of Tolerance and the logicality criterion are seen to be two inextricably entwined aspects of a solution to a fundamental problem that Carnap searches a solution to and which characterizes his aspirations throughout the period under consideration here, i.e. the problem of the rationality of scientific discourse under the variability of linguistic systems of knowledge representation.

I depict the overall development of Carnap's philosophy with this central idea continually in focus. As a supplement to the interpretation of Carnap's program as a concerted attempt to look for the fundamental invariants of thought and experience, I provide the view that a necessary condition for implementing his ideal of explication is a coherent formulation of what might be called the task of providing *genealogies* of important scientific concepts and ideas. This complies with the attractive account represented by Howard Stein about the two basic functions of philosophy, i.e., a distinction between "the enterprise of knowledge" and the "enterprise of understanding" [Stein (2004)] It is argued here that an essential ingredient of Carnap's method of explication is a variety of philosophical history of science which provides the necessary insight into the problem complex one is tackling with under the purview of explication. Therefore, a significant role is bestowed upon historical knowledge and historiography. I attempt to accommodate this aspect of the "enterprise of understanding" within the more explicitly confined "enterprise of knowledge" that Carnap was overtly concerned with. However, it is argued that the "enterprise of understanding" constituted an equally important aspect of Carnap's philosophical program, although it remained covert in his publications.

Having provided a critical assessment of Carnap's program of explication, we indicate its ramifications within the domain of values. In a certain sense, then, although Carnap's philosophical program did not manage to articulate the elements that on the basis of the general and systematic discussion conducted in this dissertation can be considered as essential for a scholar's vocation, we can see Carnap as a rare exemplar of the scholar's vocation. His lasting value for philosophy is not reduced by the fact he did not articulate a systematic theory of ethics. It is only that his philosophical significance would have been of even higher order of magnitude had he taken seriously the possibility of constructing ethics within the domain of rational discourse. With the hindsight provided by Carnap's example we can finally formulate in a general level the challenges faced in the education of new generations of scientists. Having provided us with a skeleton of a critical praxeology, Carnap has done us a tremendous service by enabling us to orient ourselves within the problem space of modern science and technological society, encouraging us to boldly encounter problems such as dealing with risk, science and survival, and ultimately, the place of science in modern life. [8]

[8] A final note on a few technicalities concerning the typography in the text. Firstly, the distinction between use and mention is marked by using double quotes for terms or phrases mentioned as in, for example, referring to the paradigmatic sentence of Aristotelean syllogistic containing metavariables: "*A* is *B*". Secondly, single quotes (also known as 'sneer quotes' or 'scare quotes') are applied to indicate special or unusual (or, if you wish, 'tongue-in-cheek') uses of terms or phrases. An example of these is the use of "good" and "bad" in the phrase "Evidently, those 'good' people have done much evil to the allies, while we, the 'bad' ones, have caused them many benefits". (Vlastos: *Socrates, Ironist and Moral Philosopher*) I do not, however, restrict my use of single quotes to sudden flights of irony. Another way of using them is marking the instances where one could not come up with a term or phrase that conveys exactly what one intends to say. In these instances, which I hope are not too numerous, I have used single quotes to accentuate literary expedients. Thirdly, italicization is used for emphasis, for established foreign (mostly German, French, Latin and transliterated Greek) terms and phrases conveying special meanings *and* for mentioning books and monographs. The dissertation was typeset with LaTeXin $\mathcal{A}_{\mathcal{M}}\mathcal{S}$ fonts using the *PCTEX* version 6.

Part I

Outline of the Problem's Background in the Intersection of Intellectual History and Systematics

Chapter 1

HISTORIOGRAPHY AND THE DIALECTIC BETWEEN AGENT-BASED AND STRUCTURAL EXPLANATIONS

> "In historic events the rule forbidding us to eat of the fruit of the Tree of Knowledge is specially applicable. Only unconscious action bears fruit, and he who plays a part in an historic event never understands its significance. If he tries to realize it his efforts are fruitless." — Tolstoy, *War and Peace*, book *XII*, ch. 2.

In this chapter I investigate some pertinent views about historiography that have figured prominently in modern European and American scholarship. I will also consider systematic issues related to the methodologies of history and sociology of science as well as issues related to conceptual, formal and mathematical tools that are expedient in studying the complex networks of influence that characterize the collaboration of scientists and the acquisition of knowledge in the framework of 'collective cognition'. The methodological inquiry undertaken in this chapter is at bottom motivated by the observation that in order to study the complementary domains of fact and value, distinct methods of enquiry are required in the two domains. The fundamental distinction between *agent-based* and *structural* explanations is here taken as a point of departure, because (i) traditional historiography and 'integrated general history'[1] (a significant part of this work may be included within the latter field) may benefit from a research orientation that has recourse to both kinds of models of explanation, (ii) circumventing the problems that result from the legacy of historicism in philosophy is possible only through a non-biased analysis of the constructive value and methodological limits of historiography; it is not possible to ignore this issue altogether, and (iii) any systematic enquiry concentrating on the domains of fact and value, and their relationship, necessarily presupposes this kind of division of labour (not in the sense of a Collingwoodian absolute presupposition but rather

[1] A term coined by Jonathan Israel.

in the sense of an ontological demarcation between natural and criteriological).[2] The various methods introduced in this chapter include (1) the re-enactment model of historical explanation of R. G. Collingwood, (2) the narrative model of historical explanation of Arthur Danto, (3) the structurist model of historical explanation by Christopher Lloyd, and (4) theoretical models of collective cognition (exemplified by David Wolpert's model of collective intelligence for representing complex interactions among a group).

Although the necessity of putting at our disposal a variety of tools as diverse as the ones mentioned above may not seem evident on first look, I can only hope that the reader can bear the suspense until chapters two and three, where the fundamental role of agent-based modes of explanation in the history of the concept of the moral is demonstrated. I take it that the relevance of the sctructurist modes of explanation is an uncontested issue in philosophy of science as the great majority of our scientific knowledge corpus is based on them. In any case, the latter play a crucial role in assessing the substantial content of the ideal of the unity of science.

1.1 Methods of Historiography

For history to make any claims about its status as a science, it must be based on a careful analysis of the possibilities regarding the extent to which the crucial processes of history, both in the socio-economic context as well as in the individual context, can be made into a coherent domain of *scientific inquiry*. It would seem that a pluralistic approach to historical research in as much as it means a recourse to relativism, post-modernism, pragmatism and 'common-sense',[3] is in a sharp contrast with a scientifc conception of historical en-

[2]This distinction is a natural (natural in the sense of appertaining to the order of things) one, i.e., one determined by the *essential form* of the phenomena within both domains. It is manifested in the essential difference between *Is* and *Ought*.

[3]The last one having, presumably, very little to do with its remote ancestor, *senso commune*. A delicate, yet important distinction is figured here. The association of 'common-sense' philosophy with the *ordinary language* philosophy of G.E. Moore and *post-Tractatus*-Wittgenstein has mainly caused the confusion. The philosophy of *senso commune* has its roots in an altogether different intellectual and cultural climate of the *Renaissance*. Indeed, the explicit definition of *senso commune* is to be found in Giambattista Vicos's *La scienza nuova*: "The *senso commune* is a judgement without reflection, experienced in common by a whole class, by a whole people, by a whole nation, or the whole mankind". ["Elements XII", *Scienza nuova*, 142] Its meaning is further clarified by Eric Voegelin [Voegelin (1998), 133–134]: "The *senso commune* is the point of origin of a civilizational course. It comprises the primordial religious and legal institutions of a nation, and the unreflected ideas embodied in these institutions are the stock of meaning that is penetrated in the historical course increasingly by reason, until at the *akme* of the course, the moment of perfect balance between substance and reason is achieved. The meaning of the *corso* is the refinement of an initial, dense, unreflected substance to a maximum of rational differentiation. The later, rational phase does not add to the substance. Reason can operate only on the initial stock. [...] Under this aspect, the concept of the *senso commune* established the great principle of civilizational interpretation that the history of a civilization is the history of the exhaustion of its initial myth and of such mythical elements as may have entered the course from other resources." It is thus an understatement to say that "the philosophy of ordinary language is not exactly the same as the philosophy of common sense, yet historically speaking they treat overlapping themes", as is related by Philip Larrey in *Sensus Communis* Vol. 5 (2004), No. 2-3 (April - September). Although the latter clause is undoubtedly true, the intended contrast is too weak. Although one could, feasibly, be sceptical

quiry. For my purposes in this dissertation, I assume that in history it is possible to adopt an attitude that is akin to *scientific realism*. I do not wish to engage here in an in-depth discussion about the notorious controversy between realism and anti-realism which forms the background for the general debate about historiography, largely instigated by the post-modern theorists.[4] From the present point of view it is sufficient to describe the realistic underpinnings of the methodology adopted here by referring to the distinction between the 'theatre of history' (history *a parte objecti*) and the interpretations of historians (history *a parte subjecti*). Both aspects figure in a realistic conception of history, in as much as history *a parte objecti* forms the ontological grounding of any meaningful scientific investigation, and in as much as the world is studied by means of the methodological tools available within the purview of history *a parte subjecti*. This amounts to a conception of history which affirms the belief that the world of human affairs is, in principle, within the power of humanity to control. Given that different historians have different goals and priorities, the ideal picture given above raises suspicion. Is there, then, any possibility of ordering the goals themselves in a way that guarantees the 'objectivity' of the realistic representation? One way of answering this question in the positive is grounded in R. G. Collingwood's conception of the scale of forms.[5] E. H. Carr provides an example of how such a standard operates in practice:

> 'Historiography' is a progressive science in the sense that it seeks to provide constantly expanding and deepening insights into a course of events which is itself progressive. [...] To take the simplest of illustrations. So long as the main goal appeared to be the organization of constitutional liberties and political rights, the historian interpreted the past in constitutional and political terms. When economic and social ends began to replace constitutional and political ends, historians turned to economic and social interpretation

about the prospects of doing 'common-sense' history, from the point of view of the rooted consciousness of social evolution embodied by Vico, the *senso commune* would form a valid basis of historical enquiry.

[4]The post-modernist scholars mainly draw on the philosophical writings of such figures as Barthes, Foucault, Deleuze, Derrida, Lyotard, Baudrillard, Irigaray, Kristeva, Lacoue-Labarthe, Spivak, Judith Butler, Laclau, Sande Cohen, Stanley Fish, and Richard Rorty, to name a few. The central tenets of the post-modern historiography can be stated as follows: (1) the past has gone forever and that our statements cannot, therefore, be said to correspond to it; and, (2) language as such is referential and characterized by endlessly deferred chains of meaning. [Connelly (2006) [Macfie (2006), 187]]. Some of the most vigorous proponents of the post-modernist program in history are Alun Munslow [(1997): *Deconstructing History*, Routledge, London and New York], Keith Jenkins [(1991): *Re-Thinking History*, Routledge, London and New York; (1995): *On "What Is History?" From Carr and Elton to Rorty and White*, Routledge, London and New York; (1997): (ed.) *The Postmodern History Reader*, Routledge, London and New York; (1999): *Why History? Ethics and Postmodernity*, Routledge, London and New York.], Robert Berkhofer [(1995): "A Point of View on Viewpoints in Historical Practice" in F.R. Ankersmit and H. Kellner (eds.) *A New Philosophy of History*, Reaktion Books, London.] and Frank Ankersmith, [Ankersmit, F.R. and H. Kellner (eds.)(1995): *A New Philosophy of History*, Reaktion Books, London.] The post-modernist position has been succinctly adduced by Michael Stanford (representing 'traditional' historians) in the following way: "[...] In historiography the representation is the reality — texts are self-referential and do not refer to anything else; to such texts only aesthetic criteria are relevant, not epistemological norms or standards; we have no established texts and no past, but only (more or less plausible) interpretations; criteria of truth and falsehood are inapplicable to historiography; historical accounts are opaque and cannot be paraphrased; the historical past is only the creation of the present historians, rather than existing in its own right — this is 'constructivism' " [Stanford (1998) 234].

[5]Collingwood's conception of history is outlined in 1.2.

> of the past. In this process, the sceptic might plausibly allege that the new interpretation is no truer than the old; each is true for its period. Nevertheless, since the preoccupation with the economic and social ends represents a broader and more advanced stage in human development than the preoccupation with political and constitutional ends, so the economic and social interpretation of history may be said to represent a more advanced stage in history than the exclusively political interpretation. The old interpretation is not rejected, but is both included and superseded in the new. [Carr (1962), 118]

The realistic picture of historical research as reaching towards more accurate depictions and more comprehensive explanations of historical phenomena functions as an ideal, a sort of Peircean limit of research, that regulates the formation of the variety of explanatory strategies and patterns presented by historians. This picture of historiography openly consents to the view that there exist severe practical difficulties of discovering and conceptualizing the relevant elements of historical explanation. History is irreducibly theory-laden, regarding equally the suggested patterns of explanation as well as the generally accepted methods of collection of data. However, this does not exclude the possibility of envisaging history as an epistemological project aspiring continually towards better and more refined explanations, the corroboration of which is the task of competent historians. The kernel of realism included in the overall view propounded in this dissertation is just the contention that there exist *historical processes* (independently of our representations), but that these become objects of *scientific investigation* only via our theoretical representations. These representations, in turn, provide us with an increased understanding. Indeed, as Collingwood put it:

> The historical process is a process in which man creates for himself this or that kind of human nature by recreating in his own thought the past to which he is heir. [. . .] [B]y understanding it historically we incorporate it into our present thought, and enable ourselves by developing and criticizing it to use that heritage for our own achievement. [Collingwood (1946); 226, 230]

In the following two sections I look forward to delve in depth into the more philosophical aspects of historical explanation. In order to tackle the main difficulties pertaining to the task at hand, I will adduce a few historiographical methodologies that have figured in philosophy in the twentieth century, mainly in the European and American traditions.[6]

[6]As is made evident by the group of representative thinkers of historiography I draw on in my study, I have no sympathies for the analytic/continental divide taken in an ideological sense. That it is a tangible distinction in a historical, sociological and topical sense I do not wish to deny (although some will deny even this!), but to choose between the representations of the analytic and the continental traditions of historiography as an attempt to discover methodologically sound principles, is in my opinion futile, not least because philosophy generally lacks a definite methodology and because the conceptual tools and styles of argumentation of different philosophical 'schools' may (and in particular cases do) overlap significantly. I am inclined to think that the approaches represented by the thinkers I am about to adduce, supplement, rather than contradict each other. It is beneficial to have various methods and modes of thought at hand;

I will assess these methodologies in the light of the most urgent priority of current historiography, viz., the necessity of redefining the field of 'history of philosophy' in any contemporary attempt to devise a methodology of intellectual history capable of serving as a frame for 'integrated general history'. The necessity of redefining the field of history of philosophy derives, essentially, from the confusing present state of historical studies in general and from severe misunderstandings in the study of 'modernity' in particular.[7] The number of different methodological approaches in historiography has grown beyond control; so much so, that it is no longer possible for a single scholar to get a coherent overall view of the field. Some kind of unification is therefore necessary. Moreover, the history of philosophy is generally so narrowly defined that it is impossible to satisfactorily accommodate the Enlightenment meaning of the term "philosophy" within its scope. This is, of course, one of the most pressing reasons to attempt to reformulate history of philosophy from our point of view. Indeed, as Jonathan Israel has put it: "[...] without a dramatic widening of the scope of 'history of philosophy', breaking in this respect with the Lockean and Humean legacy, no historian, or philosopher, can be said to engage broadly with Enlightenment ideas about revolutions and society or deal comprehensively with a principle of 'modernity' conceived as a set of values, attitudes, and ideas generated by the Radical Enlightenement." [Israel (2006), 15]

The goal of a historian of ideas is to provide fruitful interpretations of conceptual phenomena in their historical context. These interpretations should open up new possibilities for systematic developments that still retain the valid kernel of substantial insight of the earlier suggestions and ideas. The challenge is more or less to frame hypotheses which

it is like having a toolkit that contains different tools that are appropriate for various tasks. This does not, however, amount to any sort of relativism, as I will argue in the following. Indeed, I agree wholeheartedly with Ryle, when he says that: "There is no place for 'isms' in philosophy. The alleged party issues are never the important philosophic questions, and to be affiliated to a recognizable party is to be the slave of a non-philosophical prejudice in favor of a (usually non-philosophic) article of belief. To be a 'so-and-so' is to be philosophically frail. And while I am ready to confess or to be accused of such frailty, I ought no more to boast of it than to boast of astigmatism or *mal de mer*." [Ryle (1937) [Ryle (1971b), 153–154]]

[7] The reasons for this state of affairs derive, in the first place, from the lamentably common anti-historical attitude of philosophers, particularly in the anglophone countries, and, in the second place, from the lopsided Enlightenment historiography which has failed to deliver the essential difference between the so-called mainstream Enlightenment and Radical Enlightenment. Jonathan Israel has analysed the situation as follows: "[D]ue to the leanings of much recent historiography, as well as the anti-historical orientation of twentieth-century Anglo-American philosophy, the modern reader investigating the rise of 'modernity' as a system of democratic values and individual liberties in the Enlightenment encounters a bewildering and curious paradox. For the crucible in which those values originated and developed — the Radical Enlightenment — has not only, until recently, been very little studied by scholars but at the same time confronts us with a major philosophical challenge in that its prime feature is a conception of 'philosophy' (and indeed of 'revolution') from which during the course of the nineteenth and twentieth centuries western liberal thought and historiography, especially in the English-speaking world, managed to become profoundly estranged. Part of the difficulty, in contemporary Britain and America, is that philosophy's proper zone of activity has come to be so narrowly defined by the intellectual heirs of Locke and Hume that philosophy is generally conceived to be a marginal, technical discipline which neither does, nor should, affect anything very much, let alone define the whole of reality in which we live, an approach which firmly places 'philosophy' at the very opposite end of the spectrum from the Radical Enlightenment's (and indeed Marx's and Nietzsche's) conception of 'philosophy' as discussion of the human cosmic condition in its entirety, the quest for a coherent picture, the basic architecture, so to speak, of everything we know and are." [Israel (2006), 13]

skillfully balance the contrary requirements of *accuracy* and *attractiveness*.[8] An illuminating historical study results from an optimal combination of these qualities.

1.2 Agent-Based Explanations in Intellectual History

Collingwood's re-enactment model of historical explanation

"[...] there is no such thing as knowledge either of the particular or of the universal, but only of the individual [...]" — R. G. Collingwood, "Are History and Science Different Kinds of Knowledge?"

Underlying Collingwood's philosophy of history is a fundamental distinction of metaphysical origin. According to Collingwood, the function and scope of the human mind should be characterized along the lines of a dichotomy between thinking and feeling.[9] Regarding the cornucopia of activities and functions that can be accommodated under the generic term "mind", this distinction amounted to a thoroughgoing division of the entire range of these activities falling under either thinking or feeling. Moreover, not only the metaphysical contrast was drawn as a sharp demarcation, but Collingwood even went so far as to find it appropriate to speak of some actions as "determined by thought" and of others as determined by "mere impulse and appetite". [Collingwood (1993), 216] Hence, a value-theoretic position was also built into the framework of mind that laid the basis for his philosophy of history. From a methodological point of view, thought has a primary role as an explanatory element in human sciences. Indeed, those actions in which the thought of the agent plays an essential and causative role are termed by Collingwood as "reflective" or "rational" actions. Characteristic elements of reflective thinking include the agent's situational awareness (providing his motives for acting), his purposes, his scruples and values, and his dispositions, his know-how,[10] prudence (weighing of practical means and consequences), etc.

The emphasis on the rational aspects of human agency was in fact a necessary ingredient of Collingwood's conception of history. Moreover, it was an ingredient which took a concise and definite form in his thinking only gradually. He spent a long time searching for such a methodological basis that could provide him with a criterion to discern the rational content of history as an academic discipline, and enable him to justify the underlying overall vision of his according to which history and philosophy were essentially one and the same discipline. Indeed, the *rapprochement* of philosophy and history was the main theme of all of Collingwood's later writings. Collingwood crystallized his conception of the purpose of philosophy in a lecture on the philosophy of history in 1930

[8] A pair of terms used in [Carus (2007), 40n]

[9] This contrast is clearly visible in all of Collingwood's later writings, from *The Idea of History* on.

[10] In some cases, a reference to the agent's expertise could be more accurate and illuminating, comprising such procedural maxims as (1) reinvestment in learning, (2) seeking out more difficult problems, (3) tackling more complex representations of recurrent problems, etc. For a comprehensive treatment of this kind of cognitive characterization of agents, see [Bereiter & Scardamalia (1993)].

in the following terms: "Philosophy is thinking about the world as a whole. To study the nature of selected parts of the world is to be a scientist; to study its nature as a whole is to be a philosopher." [Collingwood (1930) [Collingwood (1965), 121]] To see how such a conception of philosophy fitted with Collingwood's conception of history, it is expedient to clarify how he saw the relation of philosophy to other disciplines in general. In the same lecture he maintains that " 'The philosophy of something' is a legitimate phrase only when the 'something' in question is no mere fragment of the world, but is an aspect of the world as a whole — a universal and necessary characteristic of things." [ibid., 122]

History was an ideal companion to philosophy in this respect. The purpose of history, according to Collingwood, was not encapsulated by the 'scissors-and-paste' style of history that prevailed in the academia at the time. In history it was not a question of merely providing reliable chronicles relating what had happened in the past. Even if such approaches accommodated principles that could be called 'scientific' in a very loose sense (verging on a kind of eliminative induction *á la* Francis Bacon), they did not fulfill their promise to deliver an adequate account of the past, an account that could be of use in solving problems arising at present. This was the fundamental criticism of Collingwood regarding traditional academic history. History as a science professed to be something completely different. As Collingwood put it: "If there is to be philosophy of history, history must be something more than a trade or an amusement. It must be a universal and necessary human interest, the interest in a universal and necessary aspect of the world."[*ibid.*, 123] Moreover, history is not a mere intellectual interest in its own right, it is a *form of knowledge*. It is the justification of this line of thought that becomes the *modus operandi* of Collingwood's intellectual aspirations.

It is evident that this kind of conception of history necessitated a specific view of its methodology. But Collingwood had such a view ready at hand. He was inclined to think that the peculiarity of the historical method laid in its reliance on a kind of penetration into an agent's mind. When Collingwood used such terms as "penetration" and "inside of the event", he was not alluding to some vague and mysterious ways of intuitively grasping what an historical agent thought, but metaphorically stating the aim of the science of human action: the task of human science in enquiring the actions of individual agents in their historical surroundings is to disclose the "thought" of the agent in order to explain that action and to render it rationally understood. In the case of an historian the enquiry of human actions is of course of a peculiar nature. The historian does not have direct access to the facts that pertain to the action or event under examination. He will have to rely on the usual sources of information in such cases (what is usually called 'evidence'), for example testimony, reports, documents and artifacts. On the basis of this evidence the historian reconstructs the "inside of the event" reflecting on that evidence and making inferences from it. This stage of the explanation construction might be called 'disclosing the thought of the agent'. This, however, is not the end point of research. After having constructed a plausible candidate for the "thought" of the agent that might account for the action he performed, it is necessary to ascertain that the deed performed actually is the "expression" of the reconstituted thought. Collingwood uses here the quite strong phrase "deed determined by the thought". What it essentially means is that the historian should, rather than devising arbitrary or imagined reasons for the action, consider as many *alternative* courses of action as possible, measure these against the presumable motives of the agent in question, and finally weigh all of these alternatives against the relevant available evidence. This process of *re-enactment*, although relying on empathetic identification with the agent at one stage of the inquiry, is not an intuitionistic approach to historical explanation. It aims at being a realistic one, conferring an equally important status to both inductive reconstruction and empathetic deliberation. These two aspects constitute what can be called the dimensions of *chronicle* and *narrative* in Collingwood's historical method. In his *Autobiography* Collingwood succinctly described the core of his historical methodology:

> So I reached my third proposition: 'Historical knowledge is the re-enactment of a past thought incapsulated in a context of present thoughts which, by contradicting it, confine it to a plane different from theirs.' How is one to know which of these planes is 'real' life, and which mere 'history'? By watching the way in which historical problems arise. Every historical problem ultimately arises out of 'real' life. The scissors-and-paste men think differently; they think that first of all people get into the habit of reading books, and then the books put questions into their heads. But I am not talking about scissors-and-paste history. In the kind of history that I am thinking of, the kind I have been practising all my life, historical problems arise out of practical problems. We study history in order to see more clearly into the situation in which we are called upon to act. Hence the plane on which, ultimately, all problems arise is the plane of 'real' life: that to which they are referred to for their solution is history. [Collingwood (1978), 114]

There is, however, a flavour of relativism in Collingwood's approach that has not satisfied all theorists of historiography. For example, Rex Martin remarks that

> What is disconcerting about Collingwood's suggested criterion is its notion of radical subjectivism. If there can be no appeal beyond the investigator's subjective appraisal, then we have subtly shifted the focus of our account of explanation from the apparently objective connection of deed with thought, which is what was originally meant by the statement that this deed *makes sense* in the light of that thought, to the psychology of the investigator when he asserts a connection of plausibility. [Martin (1977), 64]

Even though Martin expresses a here a well-founded worry, I think he has not adequately seen the implications of Collingwood's methodological thought in their overall context. Collingwood was not suggesting that the capricious elements of an investigator's psychology should be let to determine the content of an historical account by themselves. In fact, Collingwood was deeply aware of the risk and constantly searching for criteria that would render an intentionalistic approach scientifically acceptable.

Now, without going into the details of the criteria that Collingwood himself provided for the epistemological grounding of re-enactment, I want to make a couple of points which make his approach more plausible. In the first place, an agent's[11] action may belong to one of these ranks: (i) *impulse actions* that are closely related to reflexes, (ii) actions that are *not goal-directed* and (iii) actions in the proper sense of the term which are characterized by the conscious anticipation of results, in short, a *concrete goal* that is subjectively intended. Generally speaking, all of these might be relevant for explaining a particular historical event. However, it is only actions in the proper sense (iii) that can be satisfactorily embedded in the re-enactment scheme of explanation. The reason is that for Collingwood only actions in the sense (iii) can be subsumed under the criteriological framework which constitutes the general conditions of possibility of the process of re-enactment; a necessary condition for discerning the possible conscious motives of an act of a particular agent is that the act was intended with a particular goal in mind in the first place. An unconscious act, even if it could be *ex post facto* shown to be a result of definite ratiomorphic processes embedded in the organic functions of the agent, could not be understood within the framework of re-enactment which presupposes an intentional basis for the act under scrutiny. What is it then, that makes it possible to reconstruct the conscious motives of historical agents on the basis of historical evidence (testimony, reports, documents and artifacts)? Moreover, what makes such a view plausible? The answer lies in the conception of self and its irreducible historical character. From the point of view of a self, the accounts it provides – *for others as well as for itself* – of the cognitive, emotive and conative aspects of its psychological states as well as of the acts it takes are irreducibly historical: "knowing oneself is historical — It is only by historical thinking that I can discover what I thought ten years ago, by re-reading what I then thought, or what I thought five minutes ago, by reflecting on an action that I then did, which surprised me when I realized what I had done. In this sense all knowledge of mind is historical." [Collingwood (1936), 19] From a first-person perspective it is natural to describe the determining conditions of self-identity in such historical terms. But then there is nonetheless a conspicuous rupture between the *subjective*, experiential quality of historicity characterizing self-formation from this first-person perspective and the *objective*, purely descriptive elements of historicity to be found within scientific accounts of the development of man in the domain of historical psychology (the history of human cognition), as exemplified in the interesting work of A.R. Luria and L. Vygotsky, for example. What is noteworthy in their work is their emphasis on the socially constituted structure of self which is a product of historical evolution:

> The behaviour of contemporary civilised man is the product not only of biological evolution or childhood development; it is also the product of historical development. *In the process of man's historical development*, external relations between people, and relations between mankind and nature are not all that has changed and developed. *Man himself changed and developed; human nature has changed.* [Luria & Vygotsky (1992), 41][12]

[11] Let us not be deceived by the seeming value-neutrality of the term; we are here mainly concerned with men and it is a determining condition of their nature that the background framework of their actions include irreducible axiological (Gk. ἀξία (worth)) elements, both moral and prudential.

[12] Italics added.

They make the indisputable observation that the project of historical psychology is at the outset fraught with severe methodological difficulties, not least because the available evidence for a reliable and systematic reconstruction of the psychological structures and lawlike regularities of the self is very scarce. Especially the hypothetical constructions on the basis of archaeological artefacts are suspect in as much as they fail to give a "remotely objective or complete account of the psychological mechanisms of behaviour." [*ibid.*] It is thus not surprising that historical psychology can draw on a very much smaller body of material than is provided for the study of external history. Luria's and Vygotsky's unit of analysis is the individual, and this is so even within the domain of historical psychology. The emphasis on the individual aspects of the history of cognition immediately raises the question about the possibility of acquiring theoretically relevant information on the basis of historical evidence the partition of which into the classes of relatively numerous but categorically inadequate data (such as typical archaeological artefacts) and scarce, but intrinsically relevant data (such as protocols of different tasks involving various cognitive functions such as attention, memory, producing and understanding language, solving problems, and making decisions), seems to set unassailable limits to the scope of historical psychology.[13] Furthermore, Luria and Vygotsky are engaged with the history of cognition at two levels, (i) on the 'paleo-psychological' level of physical environment and instincts and (ii) on the 'historical' level of intellectual abstraction and representation. Their fundamental problem is to link these two levels together within the context of historical psychology. How can this be achieved? Can Collingwood's method of re-enactment serve as a mediating framework for the transition from the instinctive domain of the natural and empirical to the criteriological domain of the abstract and normative? This is a pertinent question, on the answer of which depends to a prominent degree the usefulness of Collingwood's method as an instrument of discovery.

Let us go back to the basic ingredients of Collingwood's approach. He says that historical questions are questions in which one tries to understand what somebody was doing on certain occasion. A necessary condition of this is that one understands what sort of occasion it was: "for every action arises out of the situation in which it is done, and there is no understanding the action unless one understands the situation." [Collingwood (1940), 191][14] The fundamental criterion of success, so Collingwood insists, "is to study the background". [*ibid.*] The investigator strives to embed the actions which are the object of his study within a particular situation. The stance towards the thoughts determining the act is therefore situated rather than mentalist. The ultimate ground of the re-enactment is a contextual evaluation of the background in which the act and the thought which determined it (the intentional act) took place. Reconstructing the concrete presuppo-

[13]As an example of the level of specificity at which Luria and Vygotsky elaborate on the history of the development of different cognitive skills, consider their account of the development of memory: "The historical development of memory begins from the point at which man first shifts from using memory, as a natural force, to dominating it. This dominion, like any dominion over a natural force, simply means that at certain stage of his development man accumulates sufficient experience – in this case psychological experience – and sufficient knowledge of the laws governing operations of memory, and then shifts to the actual use of those laws. This process of accumulation of psychological experience leading to control of behaviour should not be viewed as a process of conscious experience, the deliberate accumulation of knowledge and theoretical research. This experience should be called "naïve psychology". [Luria & Vygotsky(1992), 56]

[14]Italics added.

sitions of acts enables one to determine elements that constitute intentions underlying conscious acts in particular situations. One can go further and look for the presuppositions of presuppositions, and presuppositions of presuppositions of presuppositions. Does this sequence terminate? According to Collingwood, it does. The terminal points of these sequences he calls *absolute presuppositions*. The radical insight of Collingwood is to conceive metaphysics as the study of absolute presuppositions. Its method is essentially historical, as was clearly stated by Collingwood: "metaphysics can set its house in order by living up to its proper character as an historical science." [Collingwood (1940), 81] The idea was, as is well known, fiercely attacked by A.J. Ayer who, inspired by the anti-metaphysical spirit of the Vienna Circle, attempted to devise a put-down argument against the possibility of metaphysics. The argument Ayer presents in his classic *Language, Truth and Logic* effectively takes the form of the following syllogism:

(1) *Any proposition which cannot be verified by appeal to observed facts is a pseudo-proposition.*

(2) *Metaphysical propositions cannot be verified by appeal to observed facts*

$\therefore$ *Metaphysical propositions are pseudo-propositions, and therefore nonsense.*

What makes the whole account problematic, according to Collingwood, is the error of mistaking suppositions for propositions, essentially an error in logic, having its root in the way of thinking that logical efficacy, or the pragmatic feature of causing questions to arise, belongs exclusively to propositions. "Mr. Ayer, true to the positivistic tradition, does not possess the idea of supposing, and *a fortiori* not the idea of an absolute presupposition." [*ibid.*] Collingwood finishes his description of Ayer's main error with characteristically cogent statement "Any statement of an absolute presupposition which he encounters in the course of his reading, therefore, he regards as a statement of a proposition; and for the metaphysical question 'Was this presupposition made on a certain occasion or not?' he substitutes the pseudo-metaphysical question 'Is this proposition true?' " [*ibid.*] One conspicuous feature of the positivists' rejection of metaphysics is their thesis[15] that metaphysics might somehow be 'malicious' towards science (an expression of Santayana), and consequently the fear that unless the unruly entrophication of metaphysics is impeded, and the whole discipline ultimately dispensed with, it will destroy science. This thought is equally seen to rest on a mistake. Indeed, there is no controversy between metaphysics, properly conceived, and natural science, as Collingwood emphasizes: "No one is trying, or has tried within living memory, to tyrannize over natural science in the name of metaphysics. From metaphysics properly so called, the attempt to ascertain the absolute presuppositions of thought, a natural science that does its work conscientiously can have nothing to fear." [*ibid.*, 168] The whole issue is, in the end, decided with brittle sarcasm:

[15] One should not ignore the possibility of its having its roots in ratiomorphic processes.

> But this positivistic terror that metaphysics may injure natural science is, after all, not without a sort of lunatic foundation. If metaphysics were what the positivists mistake it for, if it were an attempt to provide empirical justification for the presuppositions of science, it might certainly, though without malice aforethought, prove detrimental to science itself, not by its success but by its inevitable failure; for when the discovery was made that no justification of this kind is to be had, the positivistic belief that it is nevertheless necessary might lead to the false conclusion that the whole fabric of scientific though is rotten at the core. Thus understood, we may think of the positivists as, in a way, right to fear metaphysics as he does; but what he fears is not metaphysics as it really is, but metaphysics as he misconceives it; and further, what he fears is not this phantom itself, but the frightful consequences which, as he falsely imagines, would ensue upon its failure to do what he thinks to be its proper work: a work which in fact does not need doing and cannot be done. Such fears are a proper subject of study to the psycho-pathologist. [*ibid.*, 169]

Of course, the positivists were not alone with their one-sided, even prejudiced, conception about metaphysics. The markedly anti-metaphysical attitude was an axis around which a prominent part of the scholarly philosophy in England had revolved from the time of Hobbes on. The antimetaphysical and naturalist[16] spirit has been especially linked with the tradition of natural philosophy that stems from the work of Isaac Newton, Edmond Halley, Robert Hooke and Robert Boyle. The positivist philosophies that emerged in the nineteenth century are direct descendants of this tradition in natural philosophy. Elements of this basic attitude were then mediated to the logical positivists by Ernst Mach and Richard Avenarius on the continent as well as by Bertrand Russell and G.E Moore in Britain. But even such a deep, and occasionally mystical, thinker as Alfred North Whitehead who later became known for his metaphysical writings (especially for his 'process philosophy' expounded in *Process and Reality*) remarked once that "having recourse to metaphysics is like throwing a match in a powder magazine — it blows out the whole arena." [Whitehead (1920)] The specifically anti-metaphysical attitude was thus something that was very much a part of English intellectual culture. Against this background, Collingwood's detailed analysis of the significance of metaphysics and the scope of its applications is a remarkable and rare instance of intellectual courage. Indeed, conceived as an inquiry about the conditions of possibility of a theory of absolute presuppositions, Collingwood's work constitutes one of the most important contributions of British philosophy in the twentieth century along the achievements of Russell and Moore. Furthermore, it provides for the historian of any discipline, and the philosopher in particular, the possibility of dispensing with the peculiarly cross-sectional and timeless accounts of phenomena which in reality engage persons with a history and a historically

[16]Naturalism, as intended here, refers to an historical intellectual orientation dating back to the Renaissance which is very much different from what is referred to as naturalism in contemporary philosophy. The former is the view that areas that had traditionally been believed to require supernatural explanation should in fact be subsumed under natural explanation. But even this variant of naturalism is taken with the understanding of requiring that certain capacities and powers are incorporated into nature which modern forms of naturalism have been quick to dispense with. Modern variants of naturalism resemble more closely the reductionist forms of corpuscularianism that emerged in the seventeenth century with Hobbes, for example.

developed sense of self and morality. Thus, Collingwod's method can be extremely helpful for an axiological assessment of human action in general.

Historical explanations as narratives

It is well known that both Croce and Collingwood were severely critical of the earlier theorists' attempts to model historical explanations after the explanatory models used in the natural sciences. The idea that there might be universal laws governing the course of historical events they both regarded as a gross error. Nevertheless, with Carl G. Hempel's deductive-nomological, or "covering law", model of explanation introduced in the late 1940s, there emerged a concrete tool for constructing even historical explanations on the model of lawlike regularities, a method that had proven so successful in the study of inanimate nature. The apology for the D-N model of explanation in history pleaded to the claim that explaining a given historical occurrence in terms of some other event or set of events necessarily involves an appeal – which need not be more than tacit – to *laws* or *general propositions* correlating events of the type to be explained with those of the kind cited as its causes or conditions. The debate concerning the proper methodology of historiography reached its apogee in the dichotomy between *erklären* and *verstehen*, on the former side of which were historians and philosophers sceptical about the existence of a specific method for investigating human culture and on the latter side of which were historians and philosophers advocating a complete autonomy for definite categories and procedures for human studies.

In as much as this methodological debate has had any effect on the historiographical speculations of philosophers, it has mainly resulted in skepticism about the existence of lawlike regularities. Against this background it would indeed be courageous to try to defend the idea of universal historical laws within historiography. Let us see to what extent the idea of lawlike historical regularities is plausible and how one could accommodate the idea of historical laws within one's conception of historiography.

Without delving into the many dangers and pitfalls of historical explanations[17] we should concentrate on the most perspicuous concern of any historian, viz.: the role of narratives. In as much as the majority of historical texts are written in a form of a narrative, narratives should be seen as constituting a particular form of *historical explanation*. Now, what kind of support can be lent for such a view? If narratives are in effect *stories*, what kind of conditions must be imposed on them in order to render them as instances of explanation?

An important feature of all historical explanations is that they concern a *change* (in an individual or in a situation, generally). A natural restriction pertaining to such explanations is that they should concern a given, singular individual or situation. Moreover, they

[17] Among these pitfalls could be mentioned the use of temporal language and related temporal scepticism, evidence and historical relativism, the dichotomy between history vs. chronicle, the meaning and epistemic status of narrative sentences, the future – and past – contingencies, as well as the problem of general laws in history.

should pertain to individuals (or situations, structures, etc.) for which clear and distinct criteria of *unity of identity* can be stated. These are *ontological* criteria for the identification of which strict and procedural epistemic rules have to be provided. Quite generally, it is natural to require of historical explanations (or any explanations, for that matter) that they conform to the principle of the *unity of the subject*.[18] There is an immensely challenging problem here concerning historical ontology, namely, the one of deciding what are the elements which persist through a change. To make a decision with regard to a single individual might be unproblematic (historically speaking), but once we start to consider wider and more complex historical agencies, such as feudalism, nationalism etc., we immediately realize the difficulty of this decision problem. It is therefore for the sake of mere practical feasibility that we have to satisfy ourselves with the task of explicating a model of historical explanation in the form which accommodates only individuals. Now, consider the following example of a concrete historical event which might interest historians:

> By September 1941, the international situation looked quite bleak for the world but quite positive for Germany. As is noted in the play [*Copenhagen* written by Michael Frayn], by this time the Reich had reached its greatest extent. Most of continental Europe was under Nazi occupation, German panzer divisions were plunging into Soviet Russia, and the United States was still officially neutral. Heisenberg had learned from his German coworkers that an atomic bomb was not just a theoretical possibility but that it could indeed become a practical reality. Whether the war ended with the German army in place, or bogged down in a protracted conflict reminiscent of World War I, it was easy to suppose that the United States would have enough time and resources to catch up with German researchers and build a nuclear weapon, which they might well use or threaten to use on Germany. In a memoir, Elisabeth Heisenberg wrote that throughout the war her husband "constantly tortured himself" with the thought that the better equipped Allies might build the bomb and use it on Germany.
>
> At the same time, Heisenberg probably knew or strongly suspected that Bohr was in contact with Allied scientists through underground sources. So, what was Heisenberg trying to tell Bohr during this meeting, and what did he want from Bohr? The broader historical setting and a fuller appreciation of Heisenberg's outlook and relationship to the war and to fission research strongly suggest that he wanted to convince Bohr that the seemingly inevitable German victory would not be so bad for Europe after all. The alternative, as Heisenberg later noted to his horrified Dutch colleagues, was a Europe ruled by the Soviet Union. Having witnessed a traumatic Soviet revolution in Bavaria as a teenager, Heisenberg always considered Soviet domination an

[18] This is, at least, a position that we take here for the purposes of analysis. We could envisage explanations that accommodate more general and complex explananda that feature multi-agent constituents, such as collectives (featuring collective intentionality), structures, cultures, etc. *Prima facie*, the criteria of unity of identity for such entities would considerably differ from the ones for individuals. The essential idea of the explanatory schema remains, however, the same.

> even worse evil than Nazi domination. What he apparently wanted from Bohr was for Bohr to use his influence to prevent Allied scientists, who were surely far behind the Germans, from working toward building a bomb that could be used against Germany. [Cassidy (2000), 32]

Given the historical significance of the events alluded to, several pertinent questions arise. Let us concentrate on the question: what was Heisenberg's motivation for meeting Bohr in 1941? From the point of view of historical interest, the most important presupposition of this question is that before 1941 there did not exist for Heisenberg *the* kind of motive for meeting Bohr that suddenly emerged in 1941. What brought about this change? Using this event as an exemplar of a historical explanation pertaining to a change (here a change in Heisenberg's motivation), we may represent the form of an *explanandum* in historical explanation schemes as follows (I draw here on the systematic exposition of [Danto (1965)]):

E: x is F at t_1 and x is G at t_3.[19] [Cf. *ibid.*, 236]

An essential part in devising an explanation, according to this account, is looking for a mediating factor, an event, the reference to which is required by the explanation. The missing component in the schema above is a symbolic presentation of a unique event, a presentation that encodes the occasion of *something happening to x at t_2*. This event may be of any degree of complexity as long as it functions as a relevant factor (causal or intentional) in the explanation (effectively bringing about the change in x). One kind of model of historical explanation could then be formulated as follows:

(S)

(1) x is F at t_1.

(2) H happens to x at t_2.

(3) x is G at t_3. [*ibid.*]

Now, the search for a mediating element is essentially a matter of historical investigation. We cannot have any *a priori* reasons for deciding which particular event is to account for the change that we want to explain. For example, if the explanatory model that we prefer to apply in the case of the Werner Heisenberg changing his mind is of the type above, we could not decide out of hand, *without recourse to historical facts*, which events or features in the surroundings of Heisenberg account for him changing his mind. To make it explicit, in the case of this example we could present an explanation of the following kind:

[19]Thus the explanation situation given in the example quoted above can be represented as: "E: Heisenberg is indifferent to (or perhaps skeptical about) the political significance of the Allied fission research at t_1 and $\underline{E}$: Heisenberg sees the immense political significance of the Allied fission research at t_3". [*ibid.*]

(1) Heisenberg is indifferent to (or perhaps skeptical about) the political significance of the Allied fission research at t_1 .

(2) Heisenberg learns at t_2 from his German coworkers that an atomic bomb is not just a theoretical possibility but that it could indeed become a practical reality. Heisenberg concedes that whether the war ended with the German army in place, or bogged down in a protracted conflict reminiscent of World War I, it is easy to suppose that the United States would have enough time and resources to catch up with German researchers and build a nuclear weapon, which they might well use or threaten to use on Germany.

(3) Heisenberg sees the immense political significance of the Allied fission research at t_3.

This example is artificial, of course. Nevertheless, it captures some of the features of the alleged mediating factor that accounts for the change in the Heisenberg's mind. Other possibilities might as well have been imagined, and they could equally well have accounted for the required transition in the state of mind of Heisenberg. What this situation illustrates is the singularity of the explanatory instantiations of the general schema above. To reiterate, we cannot on the basis of the general schema alone predict what should happen in the problem situation characterized by the clauses (1) and (3). Historical investigation is necessary to account for the item (2).

Apart from these considerations, the schema draws our attention to another important feature: it is already in a form of a narrative, or a *story*, that has a beginning (1), a middle (2), and an end (3). It is this general form of the explanations we might subscribe to that justifies the talk about narratives as explanatory vehicles in history. Danto, who has advocated this schema as a model of historical explanation even goes so far as to allude to a certain resemblance between this model and the alleged dialectical pattern propounded by Hegel which permeates the structure of history and is exhibited everywhere in the historical evolution of societies at large.[20] It is possible even to confer a valid sense to the Hegelian claim that the thesis in a certain sense 'contains' the antithesis and synthesis. A detailed development of this analogy and the assessment how far it could be pressed have to be left for a different occasion, however. There is a clear demarcation between the forms of deductive and historical explanatory schemes, respectively, as Danto has rightly emphasized. In the first place, what historians are usually interested in are *temporal wholes* consisting of a variety of changes of the kind exhibited above. This means that the ultimate goal of historical explanation is to give an account of a complex event having a composite structure the parts of which could be treated according to the general schema (S). To illustrate this, we have to consider Heisenberg's case from a wider point of view. Using the metaphor of the uncertainty principle, we can narrow the breadth of uncertainty regarding the seemingly mysterious circumstances of Heisenberg's visit

[20] A dialectial explanation having in general a tripartite structure consisting of a thesis, antithesis and synthesis. (This common picture perpetuated in general histories of philosophy, as has been argued by some Hegel scholars (e.g. [Forster (1993)] and [Beiser (2005)]), is misleading, or oversimplifying. Notwithstanding these criticisms, the common picture seems to me to be destitute of any significant information value, to say the least.)

to Copenhagen by expanding the focus of the historical spotlight concentrated on the singular event alone.

Indeed, the change in Heisenberg's attitude – a story in itself – is part of the middle in a larger story: the story of the progress of "Jewish physics" in Germany, a necessary condition of the spectacular development of nuclear physics and fission research.[21] This, in turn, is part of the middle of a yet larger story, the change in the attitudes of the patriotic non-Jewish Germans among artistic, academic, and military circles who wanted Germany to win the war (without wanting Hitler or the Nazi regime to win). This, in turn, is part of the middle in a further story, the change in the prospects of German culture surviving until better times, within the story of the change in United States's position towards the Nazi occupation in the continental Europe, and this in another story, the change in the power relations between European states and United States of America . . . and so on. [Cf. *ibid.*, 241]

These different explanatory tasks at various levels of complexity form a nested sequence that could be illustrated graphically as follows:

$$\begin{gathered} () \\ (()) \\ ((())) \\ (((()))) \\ \cdots ((((())))) \cdots \end{gathered}$$

This, however is an oversimplified view of the general situation. More complicated structures of mutual dependence/independence might figure in more realistic cases, such as

$$(()()())$$

which is intended to illustrate the possibility of multiple causation or overdetermination [*ibid.*], and cases like

$$(()(())())$$

where some kind of overlaps are featured. [*ibid.*] As these examples appertain to questions about causality which from the point of view of illustrating the *form* of historical explanations is a secondary problem, I will revert to the primary question of analysing the differences between the deductive-nomological explanations of Hempel and the historical or narrative explanations of Danto. This is the most crucial question regarding the

[21] As Cassidy explains, it "provided a splendid opportunity for the scientists to gain the protection of the German army through a sustained project that might produce a powerful new weapon for the German arsenal – or at least a new and plentiful source of energy to power German ships and the German economy." [*ibid.*, 31] Heisenberg later put it as follows: "The official slogan of the government was: 'We must make use of physics for warfare.' We turned it around for our slogan: 'We must make use of warfare for physics.' "

model of Danto who also regards the problem of causality as secondary. In as much as it is a useful concept at all, I take it that is is not an insurmountable obstacle for historical explanations in general.

To compare the accounts of Hempel and Danto it is expedient to consider the most simple case of explanation in the Hempelian scheme, namely the one containing an inference utilizing the rule of *modus ponens*. In this we follow the exposition of Danto, who regards this procedure as the most fruitful in distinguishing the conceptual differences between the models. In effect we attempt to make explicit the apparent excluders[22] of the scheme of explanation provided by Danto to see whether it could be developed into a coherent model of historical explanation, and to see whether the excluders themselves can be circumvented by devising such a theory.

Consider the simplest case that satisfies Hempel's criteria for explanation, namely the following scheme:

(1) $(x)(F(x) \supset G(x))$,
(2) Fa,
(3) Ga.

In this scheme (3) is the *explanandum* (effectively a sentence describing a singular occurrence containing a reference to a single individual "a") and (1) and (2) jointly constitute the *explanans* ((1) being a general law and (2) an initial condition). This account clearly complies with the criteria of Hempel, (3) being thought to follow from (1) and (2) with (and *only* with) logical necessity. Danto, however, is dissatisfied with the given form of Hempelian account with regard to at least one crucial point, namely, that the transition "$Fa - Ga$" is a change, and that it is this change that we most often want an explanation for. The main point is that these changes might not always be governed by general laws, although some combination of a change of the prescribed type and an assigned cause for the change can be so governed. Furthermore, the scheme does not admit of the type of explanations that contain different predicates in the premisses and in the conclusion. Such an inference would be generally invalid. But in history we might easily conceive of situations where the explanation does not fully conform to the simplified Hempelian scheme but which, nevertheless, are considered as explanations. For example, consider a case where we insert an additional constituent to the conclusion (3), such as "Ha" and take the conjunction of "$Ga \& Ha$" as the new conclusion. This conclusion does not anymore follow logically from the premisses, but a scheme where this conclusion is substituted for

[22] A term coined by Robert Nozick in his *Philosophical Explanations* [Nozick (1981)], where the following definition is given: "The form of these [philosophical] questions is: how is one thing possible, given (or supposing) certain other things? Some statements $r_1, \ldots r_n$ are assumed or accepted or taken for granted, and there is a tension between these statements and another statement p; they appear to exclude p's holding true. Let us term the r_is apparent excluders (of p). Since the statement p also is accepted, we face the question of how p is possible, given its apparent excluders. [*] Note that the question is not: given p, how are the apparent excluders possible? Tension and incompatibility are symmetrical relations among statements, yet typically philosophical problems focus on the possibility only of some statements on one side of the relation. It is an interesting issue, what determines in which direction the question is salient." [*ibid.*, 9]

"*Ga*" it could still represent a perfectly relevant explanation in history. Another example is provided by the case where we add a constituent consisting of a monadic predicate as a new premiss, e.g. "*Ea*". From a logical point of view such a premiss is redundant, playing no role in the deductive inference. But historical explanations often contain such superfluous and deductively inert components. Although an ideal historical explanation might be pruned from such elements by the principle of paucity, there are no *a priori* reasons for considering them as detrimental to the explanatory force of the explanation as a whole. Indeed, many narrative accounts contain such inert elements, and their function might be conceived as enhancing their readability and comprehensibility.

What about narrative unity? If the logical considerations above give some clue as to the extent and nature of information contained in a historical explanation, what can be said of the underlying ontological requirements concerning the identifiability of the individual that the explanation is about? More precisely, what are the necessary conditions for narrative unity? Danto gives the following criteria:" If N is a narrative, then N lacks unity unless (A) N is about the same subject, (B) N adequately explains the change in that subject which is covered by the explanandum, and (C) N contains only so much information as is required by (B) and no more." [Danto (1984), 251] These criteria should be scrutinized carefully, especially the criterion (C): it is possible to devise a set of other criteria that explicitly conflict with (C). It in no way contributes towards the obscurity of a narrative if it contains some carefully selected 'extraneous' material. Although Danto does not deny this, he is not in a position to tackle this question in full as his main focus is on the question of the rôle of narratives. Indeed, an uncharted territory lies here for philosophers interested in the pragmatic aspects of formulating historical explanations.

So far we have only illustrated the most simple case of a narrative, a form that Danto calls an *atomic narrative* having the structure (S). The change depicted by such a scheme may not be governed by a general law, but once an allusion to a *causal* episode H[23] is made, then some general law is appealed to.[24] The general structure of such an atomic narrative is depicted graphically as follows [Danto (1984), 251]:

$$\begin{array}{ccc} F & & G \\ / & \cdot & / \end{array}$$

In this presentation the strokes represent the termini of a change while the dot represents the cause of the change. Moreover, historians usually have recourse to more complicated narratives than this, i.e. they have to devise what Danto calls *molecular narratives* of the form [*ibid.*, 252]:

[23]The symbols for general episodes are used autonomously, i.e. we do not *label* the objects of the language (the 'episode-language') and their names separately.

[24]For instance, in accordance with the simplified scheme above, this kind of law could be tantamount to a statement that H-like things cause a-like things to change from F to G.

$$\begin{array}{ccccccc} F & & G & & H & & I \\ / & \cdot & / & \cdot & / & \cdot & / \end{array}$$

In this scheme each unit $/ \cdot /$ is covered by a general law, but there need not exist a law that covers the entire range of the full process of change from F to I. It might be argued that recourse to the general scheme of molecular narratives is essential for historians (in addition to the atomic narrative form), because historians are, on the whole, ultimately interested in the large change $F - I$ that is constituted by the atomic narratives $F - G$, $G - H$, etc. It is just because the actual subject of interest is the entire process of change that we can not cling to the atomic case, for example the atomic part $H - I$ in the explanation. Moreover, the analogy between deductive arguments and narratives breaks down even at the level of form. Supposing we could model the transition from F to H by a Hempelian scheme, we should introduce a new law of the form $(x)(G(x) \supset H(x))$ which, indeed, accounts for the overall change together with the law $(x)(F(x) \supset G(x))$. But this amounts to the following discrepancy. It is a valid principle of inference in first-order predicate logic that from $(x)F(x) \supset G(x)$ and $(x)(G(x) \supset H(x))$ we can deduce $(x)(F(x) \supset H(x))$. But then the original two laws collapse into one, namely $(x)(F(x) \supset H(x))$. This kind of an elimination is not, however, always possible in valid narratives, as Danto correctly enunciates. [*ibid.*, 253] The upshot of this examination is that a valid narrative often requires more than one cause to account for a large-scale change.

> Indeed, we might even have laws in which the required initial conditions are to be satisfied in sequence, for example, $(x)(F(x)_{t_1}.G(x)_{t_2}. \supset .H(x))$, where the subscripts indicate the order in which the initial conditions are to be satisfied. We might term such laws *historical* laws. Then, with the aid of historical laws, together with a specification of temporally distinct initial conditions, we could indeed deduce our conclusion. Such laws would, in fact, enable us to make predictions, or better, *qualified* predictions. For since the two forms $(p \cdot q) \supset r$ and $p \supset (q \supset r)$ are demonstrably equivalent, it follows that we have an *historical law* of the form
>
> $$\left(C^0_{t_0}.C^1_{t_1}.\ldots C^n_{t_n}\right) \supset E$$
>
> and, if C^0 occurs at t_0, we can predict that E will take place *if* $C^1, \ldots C^n$ take place in the required temporal order. [*ibid.*, 254][25]

On the basis of this analysis, there could very well be historical laws. The existence of such laws does not, however, exclude the possibility of a compatibilist basic attitude towards the lawlike features of the world as against a fully deterministic outlook of reality. Indeed, Danto correctly notes that the existence of such laws would not entitle us to conclude that there is historical inevitability any more than the existence of natural

[25]Italics in the original.

laws entitles us to entertain the view that there is inevitability in nature.[26] There remain, then, some fundamental problems concerning the nature of narratives and general laws in history. In the first place, it could be contested whether we can transform every molecular narrative into a deductive argument. Even if *some* historical laws were discovered, it would remain uncertain whether such laws could be applied to each molecular narrative, or answering this question in the negative, whether for each such narrative a general historical law might be found. [*ibid.*] In the second place, it might be claimed that the explanatory force of narratives is weakened by such considerations pertaining to general laws. This, however, is contingent on the supposition that the recourse to general laws is a *necessary* condition of narratives. This is seen to be a mistake. It is an intrinsic feature of molecular narratives that they do not have to be based on general laws that cover the entire range of a change that is the object of explanation. It is enough that its atoms conform to this pattern. Therefore, the existence of general laws might well be *sufficient* for historical explanation, not *necessary*, and thus the absence of a general law that accommodates a whole process of change of interest does not jeopardize the explanatory task of history. To sum up, historical narratives are applied to accommodate for changes, often large-scale changes spanning vast periods of time. In Danto's words, "it is the job of history to reveal to us these changes, to organize the past into temporal wholes, and to explain these changes at the same time as they tell what happened – albeit with the aid of the sort of temporal perspective linguistically reflected in narrative sentences." [*ibid.*, 255]

1.3 Structural Explanations in History

It is a general contention that historical inquiry is necessarily based on a particular point of view determined by definite presuppositions without which the inquiry would be inconceivable. For example, in historiography it is obvious that one has to assume that there exists a past out of which the present emerged and that out of its traces in the present we can construct a picture of the past. But apart from this least common denominator there are numerous constituents comprising a point of view that may vary: (i) what constitutes *evidence* in a historical enquiry is relative to the historian and her capability of forming appropriate questions and her faculty of discernment; (ii) *principles of interpretation* are known to change over time and with the acquisition of new knowledge; also the influence of fashion should not be underestimated; (iii) organizing *categories* and *concepts* will evolve and their valuation and significance will depend on the general course of intellectual life; (iv) *explanatory theories* often change and develop along the categories and concepts related to them; they are also affected by the refinement of methodological practices used to evaluate them and by the kind of evidence that is taken to bear on them. All in all: "[...] given the multiple kinds of viewpoint and the many places they enter a historical discourse, one might accede to F. R. Ankersmit's proposition that it is only by taking a point of view that historians create in the first place historical narrative or

[26] I will come back to the issue of determinism in natural science in the subsection 4.9.1.

interpretation as such".[27]

This conclusion does not have to entail acquiescence to relativism. There is ample evidence for the thesis that it is possible to reach a satisfactory degree of commensurability among the perspectives represented by different historians. This presupposes, first of all, that there is at least a partial overlap between the *objects* of their study. Secondly, it presupposes that one is prepared to write the particular narrative or interpret the available evidence according to the interests and demands of the point of view, or perspective, in question. In setting up these criteria, a thorough analysis of the logical and epistemological aspects of situatedness (concerning, in general, situated agents and situated activities within specific contexts) is one possible option to make our historical inquiries more rigorous. Is it possible to establish the intersubjective validity of historical accounts by embedding the situatedness of agents within particular structures? Let us see to what extent the ideal of structural objectivity can be maintained within historiography by following this idea.

Structurism

To set up a framework for explicating the event/structure distinction, I will briefly consider the general methodological desiderata of structural explanations in historiography. The *structurist* conception of history that is here taken under scrutiny may be regarded as a variant of structural realism exemplifed in the general outlook that history as a science should deal principally with the problem of the history of *social structures* and not with the history of events and actions. Christopher Lloyd, one of the advocates of structurism, describes the general character of his programmatic approach as follows: "A central contention herein will be that institutionalized 'disciplines' of economic history, social history, historical political economy, and historical sociology should be considered together as one domain of enquiry — the domain of social structural history." [Lloyd (1993), 6] It should come as no surprise that advocates of structurism emphasize the inadequacy of 'common sense' to grasp the macro structures of economies and societies, as well as the causal mechanisms of their formation. Regarding the complexity of the phenomena concerned, it is nothing short of a tautology to state that "only a form of analysis and a mode of understanding that penetrates the obscured structural relations and imperatives of economies and societies can begin to reveal and explain the real history and powers of the organizational basis of social life." [*ibid.*, 1]

From a methodological point of view it is necessary for structurism to make a distinction between the study of events and the study of structures.[28] Indeed, the event/structure distinction is seen to be both ontologically and epistemologically justified, providing a rational basis for the theoretical constructs of structurism. This is reflected in the non-conservative ontic commitments regarding the fundamental entities of Lloyd's theory:

[27] Berkhofer, R. (1995): "A Point of View on Viewpoints in Historical Practice", in F.R. Ankersmit and H. Kellner (eds.) *A New Philosophy of History*, Reaktion Books, London.

[28] Structures include, according to Lloyd, at least the following entities: political systems, mentalities, and cultures as much as economic and social systems. [*ibid.*, 6]

"I shall argue that social structures (including economies) are neither patterns of events, actions, and behaviour nor reducible to social phenomena, but have *a form of structural existence* that is at once relatively autonomous but not separate from the totality of phenomena that occur within them. Nor are structures holistic or completely autonomous." [*ibid.*, 6][29] The ambitious program of structurist explanation in history attempts to vindicate the following theses:

1. There can be formulated a scientific domain of social structural history.
2. Structurism is the most appropriate basic methodology for the domain of structural history explanation.
3. Structurist methodology and structurist theory are mutually reinforcing.
4. Structurism and realism are the proper foundations for a science of structural history.
5. Structural history should be part of methodologically unified socio-historical science. [*ibid.*, 9]

In order to establish such theses it is necessary to have a clear sense of the interrelationship between philosophy, methodology and theory in scientific (historical) explanation.[30] In the first place, one should pay attention to the methodological and philosophical presuppositions of a theory. Knowledge of them enables one to gauge the relative strengths and weaknesses of alternative explanatory models; if the corresponding presuppositions have not been exposed, it becomes impossible to compare the theories and explanations entertained within a particular domain. What do these presuppositions consist in? Lloyd provides a comprehensive list of them: "Frameworks [particular aggregates of philosophical and methodological presuppositions] contain metaphysical beliefs (often well founded), general concepts about the nature of the objects of enquiry within the domain, general methodological principles, and a collection of linguistic explanatory

[29] Italics added.

[30] This criterion is admirably fulfilled in the first chapter of [Lloyd (1993)], where he distinguishes between three different problem types: "(a) *Philosophical problems*, which concern issues about existence and explanation — i.e. very general ontological and epistemological issues that remain tacit for most of the time and are only analysed by explicitly philosophical enquiry. (b) *Methodological problems*, which are more concrete in that they concern the delineation of domains and the actual explanatory practices and forms of reasoning of particular sciences and disciplines. Methodological issues have a more general currency than philosophical ones, particularly in times of crisis. (c) *Scientific theories*, which are concepts, models, and statements of a general kind about the structural mechanisms, powers, and causal relationships between types, kinds, and classes of entities, events, and processes within a domain. Theories are used directly to explain particular events and processes. We should distinguish general theories, which attempt to encompass all the main structures, mechanisms, relationships and phenomena in a whole domain (e.g., general relativity, quantum thermodynamics, plate tectonics, Darwinian evolution, neo-classical economic equilibrium theory, historical materialist class theory, Freudian psychoanalysis) from theories of more particular events and processes. The latter are formulated employing general theories, concepts, and linguistic devices such as analogies, similes, and models, but they are evaluated by bringing them into direct confrontation with empirical evidence, a procedure that may force changes in both the theory and the organization, character, and meaning of the evidence." [*ibid.*, 30]

tools, such as metaphors, analogies, similes, and source models. Scientific theories employ these beliefs, principles, tools, and models to construct putative causal explanations of types of phenomena and processes so that particular phenomena and processes can be explained." [*ibid.*, 28]

In the second place, one should establish criteria which regulate the introduction of theoretical terms of the language $\mathcal{L}$ of the theory. In other words, one should delineate the subject-matter and a body of other claims, which Dudley Shapere has referred to as "domains" and "background information".[31]

An important role in the process of domain constitution is played by what Shapere calls "relevance-relations". Lloyd interprets these relations as "structural causal relations that should be naturally delineated". [Lloyd (1993), 36] How, then, are the domain-constituting activities coördinated? In short, by the iterative procedure of devising appropriate methodologies and theories on which discoveries are dependent and which in turn disclose the ways in which the world is naturally structured. These discoveries of 'natural structure' then constitute the boundary conditions for domain construction and development. Through implementing multiple cycles of this iterative process, domains are charted which correspond ever more accurately to the part of nature or phenomena under investigation. Furthermore, "the appropriateness [of the methodology] can only be determined historically *ex post facto* and thus methodologies and theories are contingent. The methodology of the domain is dependent upon the subject-matter of the domain, for it must be appropriate for making compositional and evolutionary discoveries at the level of the particular subject-matter". [*ibid.*] However, there do not seem to occur such revolutionary transitions from 'old' methodologies to 'new' ones as Kuhn envisaged. A more moderate dynamics seems to correspond more accurately to the scientific practice:

> Incremental stages of long-term progress can be observed within sciences, with the occasional change of direction and unification of separate domains. New understanding and knowledge within science are able to incorporate and interpret older understanding and knowledge, so that contemporary theories and explanations can be seen as genetically related to earlier ones. [*ibid.*, 37]

[31] Lloyd's conception of the development of the rationality of science is constituted by the formation of these two features. Here he follows Shapere who elaborates on this theme as follows: "The development of science thus consists in a gradual discovery, sharpening, and organization of relevance-relations, and thus in a gradual separation of the objects of its investigations and what is directly relevant thereto from what is irrelevant to those investigations: a gradual demarcation, that is, of the scientific from the non-scientific. Indeed, to the extent that an area of human activity manifests the sorts of developments I have been describing, to that extent the area is considered pragmatically scientific. In other words, *this* is what we have come to *call* 'scientific'. In that development, science aims at becoming, as far as possible, autonomous, self-sufficient, in its organization, description, and treatment of its subject-matter – at becoming able to delineate its domains of investigation and the background information relevant thereto, to formulate its problems, to lay out methods of approaching those problems, to determine a range of possible solutions, and to establish criteria of what to count as an acceptable solution, *all in terms solely of the domain under consideration and the other successful and doubt-free beliefs which have been found to be relevant to that domain; that is, to make its reasoning in all respects wholly self-sufficient*." [Shapere (1984), xxii–xxiii] (Italics in the original.)

Notwithstanding considerations of the constitution of methodological principles on the basis of empirical discoveries, there is an essential tension in the methodology of the social sciences that Lloyd's overall approach to historiography aims to release. It is the tension between the explanatory roles of events vs. structures. In the social sciences, in particular, what ought to be an event/structure duality has "long and pervasively been considered a dichotomy". [*ibid.*, 44] The complementary attitudes exemplified in the attention given either to the decisions and activities of persons or to the determining powers of social entities, have given rise to two main competing explanatory methodologies – individualism and holism. These have to a large extent determined the theoretical possibilities within the field of historiography, according to Lloyd: "While these are not necessarily consciously or coherently adopted they can still be found strongly influencing the explanations of many or most practitioners in the socio-historical sciences. Fortunately, there is a third possibility – structurism – which attempts to transcend this dichotomy." [*ibid.*, 44] Without delving into the details of these three approaches, let us take a look at a summary that discloses the presuppositions determining each one of them. The following table illustrates the differences between the three approaches with respect to their ontological and methodological commitments . [*ibid.*, 47]

	Ontology	Methodology
Individualism	Only individual events and people and their actions and beliefs are real. Society is an aggregate of individuals. The term 'society' is only instrumental.	Aggregative — builds up an analysis of society by studying individuals and their motivations for action.
Holism	Society is a closed, supra-individual system with powers of self-regulation. It dominates individuals who receive their life-courses and beliefs from the whole, which acts through them.	Conceptualizes and studies the whole as a totality that structures everything within it. Searches for the internal determining mechanisms and/or essential meaning of structural evolution.
Structurism	Society is a real structure of rules, roles, relations and meanings that has to be produced, reproduced, and transformed by individuals while causally conditioning individual actions, beliefs, and intentions.	Conceptualizes and studies the structuring process over time by examining the causal interactions of individuals, groups, classes, and their structuring social conditions, beliefs, and intentions.

A summary like this does not enable us to critically evaluate the explanatory power of each approach, let alone decide which of them would be the most appropriate one for a particular need. Indeed, each can be seen to be completely dependent on the other, but for explanatory purposes one can – and it is often necessary to – use a methodology of socio-historical inquiry that emphasizes one or the other according to the needs of the explanatory task at hand. What, then, makes the structurist approach more appealing than an event-based one? Lloyd, on the basis of ontological considerations, ends up with the following diatribe for structurism:

> Structures have a superhuman, non-phenomenal existence through time, even for centuries, and they are the context and object of events, actions, behaviour, and thought. Structures can be conceived as the systems of social rules, roles, relations, and symbols in which events, actions, and thought occur and lives are lived. But structures have to be reproduced continually in thought and through action and cannot exist apart from collective thought and behaviour. The division of labour in their study should be within a methodological structurist explanatory framework that emphasizes the symbiotic duality of event/structure rather than a dichotomy. [*ibid.*, 48]

Even if the philosophical foundations of structurism are not without problems of their own, it would surely benefit historians to have an abstract overall account of the 'ideal type' of relational-structurist approach in the socio-historical studies. This is precisely what Lloyd delivers. The scheme intends to capture the features that all explanations falling under a specific designation collectively contain (or which is implicit in most of their theoretical content). According to Lloyd, "this is a pure type from which they all deviate in various particulars". [*ibid.*, 193] This ideal type consists of the following:

1. A structurist ontology and epistemology, which implies a structurist methodology, that is, explanations of any moment or part of the social totality presuppose or imply explanations of all the others. In order to explain any moment or part it must be situated in its total structural context. This is because society is a non-reducible macroscopic structure in which there is a dynamic interaction, rather than a holistic determinism, between the parts. No part is necessarily dominant over the others, but only humans have structuring power within the social structure. Structures as such do not have any autonomy.

2. A realist-relational concept of social structure. Structure is seen as relatively autonomous of individual actions and understandings but not of the structuring power of collective action over time. Structures consist of real sets of enduring social relations, rules, and roles that organize action and behaviour.

3. An abstract 'levels' model of the totality along the lines of the economy/politics/ideology/culture set of 'levels' or 'spheres' of social reality, or something similar. But the reality and the relationships between the 'levels' are major points of debate

with considerable variation in the theorization and roles assigned to these 'levels' and the hierarchical relations, if any, between them.

4. A model of persons as social agents, having self-activating powers of intentionality, rationality, reflexivity, and choice in a context of social and cultural constraint. It is people who are theorized as the makers of history but always within particular enabling and disabling social and cultural situations.

5. An important place is given to concepts of mentality and ideology. While the tradition accords a central place to systems of ideas in forming understandings of reality, it usually holds that ideas, actions and social structures can be out of phase with each other. Mentalities and ideas have to be studied for their social consequences because of their formation of understandings and motivational effects and criticized for their adequacy as articulations of social structures.

6. An important place is given to the theorization and study of social hierarchies as organizers of consciousness and loyalties, but simple class models and theories are ruled out.

7. Unintended consequences of action and unrealized results of intentions are seen as highly significant for social change. If ideas, action and structures are not mutually reinforcing then gradual social change happens irrespective of the desires of individual actors and regardless of what other forces be at work.

8. This leads to the final component — the idea that all societies are inherently changing and therefore fundamentally historical. The basic structurist idea — that society is continually being structured by agential actors, partly as a consequence of their intentions but also unintentionally, behind their own backs, as it were — is subscribed to by historians in this tradition. They therefore see the fundamental moments of the historical process, all of which have to be analysed, as being:

   a given structural and cultural circumstances that motivate, enable, and constrain action and thought;

   b action that is historically significant for its structuring consequences; and

   c the intended and unintended consequences of action that turn into the objective structural conditions that motivate, enable, and constrain action and thought, and which often appear to be unalterable. [*ibid.*, 193–194]

What is the result of these considerations? The constructive import of Lloyd's methodological approach is a unified vision of historical explanation, accommodating both the explanatory/causal role of social structures as well as individuals and singular events. The strategic possibilities of explanation are thus considerably increased. To have elucidated the importance of taking the structural point of view seriously is something for which Lloyd deserves credit. However, the full implications of the structurist view remain regrettably uncharted in Lloyd's book and an analysis of the intriguing relationship between the structural and event-based explanations is left wanting. Indeed, these questions would merit a careful treatment of their own. There seems to me to be much worthy of explication and further development in Lloyd's work, especially as regards the compelling idea of the dialectic between structural and event-based inquiries. A detailed account of such a dialectic would provide an immensely powerful tool for investigating the dynamics of knowledge acquisition and systematization within the scientific enterprise. Although the post-Kuhnian empirical inquiry of empirical inquiry has resulted in the desultory organisation of the science studies, there is no reason to believe that this domain of investigation could not be set in order through a process of clarification identical to ones that form a part and parcel of Carnap's ideal of explication. Indeed, Lloyd's overall approach would ideally suit such a task of clarification. However, concrete examples are missing from his account. A formulation of a theoretical framework in which structurist and intentionalist components work in tandem, and which provides *concrete models* of the processes of human action within structural contexts is thus a desideratum for serious historiography. With this goal in view, let us take a look at a constructive suggestion to relate the domains of intention and structure within the field of 'collective cognition'.

1.4 Collective Cognition and Modeling of the Agent/Structure Interface

"It appears that there are enormous differences of opinion as to the probability of a failure with loss of vehicle and of human life. The estimates range from roughly 1 in 100 to 1 in 100,000. The higher figures come from the working engineers, and the very low figures from management. What are the causes and consequences of this lack of agreement? Since 1 part in 100,000 would imply that one could put a Shuttle up each day for 300 years expecting to lose only one, we could properly ask 'What is the cause of management's fantastic faith in the machinery?' " – Richard Feynman, in *Rogers Commission Report*, 1986

I will concentrate here on the collective aspects of human cognition and decision making which can be regarded as an important part of a putative theoretical framework addressing the structurist and intentionalist aspects of human action in general. The study of collective cognition refers to a wide array of theoretical approaches – some of them of quite recent origin – intended to capture explanatory strategies capable of clarifying the complex phenomenon of collective action and its basis in decision processes that involve a (very) large number of individuals. It is characteristic of humans that they coördinate their actions to accomplish objectives which it would be very hard for them to attain individually. This was of course emphasized already by the ancients, e.g. by Aristotle in his well-known six-part division of ranks in *Politics*.[32] *Ranks* or *classes* are groups whose members develop common interests (in material and immaterial goods) as well as common values. Aristotle correctly observed that ranks can also be formed by groups of people who exercise no specific economic function (but for instance a military, political or religious one). The study of actions and decision processes within such ranks is of

[32] 1328b ff.

course an extremely relevant part of an integrated general history. An important step in this direction was taken during the Enlightenment when the *philosophes* realized that thinking could be seen as a collective action too – a process conducted and amplified by social groups. A demonstration of this line of thought can be found, for example, in d'Alembert's preliminary discourse to the *Encyclopédie*. Although the importance of coördinated collective action and decision making has become more and more evident towards the Modern era – the modern scientific enterprise being the most pertinent example – the theoretical explanation of their conditions of possibility and their remarkable effectiveness has remained a unattainable goal. At present, we are short of having a good theory about how collective action and cognition work. We are equally puzzled by pragmatic questions such as: when and why do collective action and cognition work, how they can be made to work better, why they fail, what they can and cannot accomplish, etc. Interesting work on these themes has been done on a wide spectrum of disciplines by investigators ranging from historians and sociologists of science to cognitive psychologists, economists and theoretical physicists. It should not come as a surprise, then, that the field is characterized by a variety of technical tools and methodological approaches. In the following I will briefly introduce one specific model of collective cognition which is relevant for the development of themes to be handled in the rest of the dissertation. The following synopsis of David Wolpert's general model of collective intelligence serves only the purpose of providing a concrete exemplar of models that appertain to the more general questions of intellectual and moral formation adduced in the following chapters of this dissertation.

1.4.1 Collective intelligence

The central question of collective cognition is: given a particular objective for group action, how does the group performance depend on the individual performances of its participants, and the ways they communicate? The traditional approach within psychology and the cognitive sciences has been dominated by the idea that the key element in the performance of a group is the isolated cognition in any individual. The (*exogenous*) effect of the group is then taken into account by assuming that the individuals are in a constant interaction with their social environment. More recent approaches to collective cognition contest this traditional 'Cartesian' approach and emphasize the importance of constant coördination and integration of cognitive and physical abilities within a group. These modern theoretical approaches are often referred to as *distributed cognition* or *extended cognition*. They are based on a variety of technical models most common of which are game-theoretical models of team or corporate action. The efficiency and explanatory power of these models is assessed by a comparison of the predicted performance of a group to the predicted individual performance. The *raison d'être* of these models is that they propose mechanisms which "explain" why co-operation can be beneficial, even in case of an extremely simple task. The rich variety of models reflects the various possible assumptions that can be made about the nature of the within-group interactions during the decision-making process. Examples of these include theoretical relationships between the parameters of members' functions (psychometric functions such as 'intelligence') and the parameters of the group function, and models of decision making. The essential criterion is that these can be tested against empirical data.

Let us now turn to the model of collective intelligence introduced by one of the leading exponents in the field, David H. Wolpert. His approach is methodological in the sense that he presents *completely general formal criteria* for any model building activity. He is especially interested in the mathematics of the design of collectives. I will now adduce the main ideas of his approach presented in [Wolpert (2003)].

Collectives

The basic constitutive idea of a collective behind Wolpert's construction is that "many systems of *self-interested* agents have an associated *performance criterion* that rates the dynamic behaviour of the overall system". [*ibid.*, 1][33] The idea is to explain theoretically how a system of self-interested agents can co-operate to achieve high "world-utility", i.e. a result that 'satisfies' both the needs of the individuals and the group as a whole. Collectives can be defined as follows:

Definition 1 ***Collectives*** $\equiv$ *Systems having the following characteristics:*

- *(i) the system must contain $N \geq 1$ agents each of which are viewed as trying to maximize an associated* ***private utility***
- *(ii) the system must have an associated* ***world utility function*** *that rates the possible behaviours of the overall system.*

In practice, collectives are often very large, distributed and support little if any centralized communication. An example is a human economy in which $N \sim 1,000,000$, the agents are human individuals and their private rewards and the world utility function can be taken to be the time average of the gross domestic product. An important presupposition of the model is that the agents avoid working at cross-purposes in order to achieve high world utility.[34] The concrete implementation of this condition is achieved by modifying agents' utility functions to be 'aligned' with the world utility. An example of such boundary conditions are the anti-trust regulations of economies designed to break up existing monopolies and prevent the formation of new monopolies. The model is such that it can be *viewed* in a way that makes it seem as if agents were trying to maximize their private utilities (this is not a built-in presupposition of the model). In a system which is viewed in this way, both the world utility and all of the private utilities are at a (local) maxima at equilibrium.

[33] Italics added.

[34] If this necessary condition is not fulfilled, phenomena like *liquidity traps* (a situation in which the short-term nominal interest rate is zero) or *Tragedy of the Commons* (depletion of a shared limited resource) may appear.

Design of collectives

Now that we have some idea of what a collective is, let us turn to the methodological issues underlying the mathematical design of collectives (more precisely, mathematical design of models of collectives). A necessary prerequisite of such modeling is to be clear about the types of problems to be solved. In his paper Wolpert presents the two central problems in the science of collectives, the latter of which is more important, and much more difficult:

The 'Forward' Problem : How the precise configuration of the system – including in particular the private utilities of the agents – affects the ensuing behaviour, and therefore affects the value of the world utility?

The Inverse Problem : How should one initialize/update the private utility functions of the individual agents so that the ensuing behaviour of the entire collective achieves large values of the provided world utility? [*ibid.*, 3]

Since the complexity of truly large systems involving a huge number of individuals precludes any detailed modeling, it is necessary to find a method to solve these problems that does not have recourse to such modeling. How can this be done? As Wolpert notes, this is a very general problem type appearing within a variety of areas: multi-agent systems (MAS's), computational economics, mechanism design, reinforcement learning, statistical mechanics, computational ecologies, (partially observable) Markov decision processes game theory. "However, none of these fields is both applicable in large problems, and directly addresses the general inverse problem, rather than a special instance of it."[*ibid.*, 3] This calls for a new, more extensive framework for addressing these problems. To many, it might seem surprising that the most popular framework of modeling collective action and decision making, viz. theory of team-games, is ruled out by Wolpert. The reason, as Wolpert points out very clearly, is that even if the subfield of game theory known as mechanism design might at first glance provide us techniques for solving the general inverse problem, it is, however, "almost exclusively concerned with collectives that are at (a suitable refinement of) Nash equilibrium. . . . That means that every agent is assumed to be performing *as well as is theoretically possible*, given the behavior of the rest of the system." [*ibid.*][35][36] How, then, is the intuitive requirement of having the private utilities of a collective 'aligned' with the world utility satisfied? In other words, how is the model to be set up, without having recourse to the theory of team-games, so that modifications an agent might make that would increase its private utility, also must

[35]Italics in the original.

[36]A detailed list of the shortcomings of the mechanism design approach includes the following: (i) it ignores all statistical issues related to how well the agents can be expected to perform for various candidate private utilities, (ii) enforced stabilization of an agent's actions will often hurt the performance of agent ρ ("How best to trade off the performance of one agent against that of other agents, or more generally, how best to trade off the degree of rationality of one agent against that of another agent?"), (iii) to have the analysis be non-trivial, restrictions like those that apply to the private utilities of human beings are needed, and (iv) it does not allow for run-time adaptive design. [*ibid.*, 4]

increase world utility? Wolpert's answer is: "Fortunately the equivalence class of such utilities extend well beyond team-game utilities." [*ibid.*, 5]

The mathematical basics of designing collectives

The mathematical theory of collectives comprises the following elements:

- individual agents are prefixed, being **players** in multi-stage non-cooperative games, with their moves at any single stage in no *a priori* way restricted by their moves at other times or by the moves of the other players
- dynamic reassignments of how the various subsets of the variables comprising the collective across all space and time are assigned to players
- modification of the player's information sets (i.e. modification of inter-player communication)
- an arbitrary vector space Z whose elements ζ give the joint move of all players in the collective in some stage
- we wish to search for the ζ that maximize the provided world utility $G(\zeta)$
- private utility functions $\{g_\eta\}$, one such function for each variable/player η
- $\hat{\eta} \equiv$ all players other than η
- a standardization of some arbitrary function U for player η is called "the *intelligence* for η at ζ with respect to U"

Intelligences are equivalent to percentiles:

$$\epsilon_{\eta,U}(\zeta) \equiv \int d\mu_{\zeta'_{\eta}}(\zeta')\Theta[U(\zeta) - U(\zeta')] \tag{1}$$

where the Heaviside function Θ is defined to equal 1 when its argument is greater than or equal to 0, and equal to 0 otherwise, and where the subscript on the (normalized) measure $d\mu$ indicates it is restricted to ζ' sharing the same non-η components as ζ. The model is not deterministic: uncertainty concerning the overall behaviour of the system is reflected in a probability distribution over Z. The ability to control the system is built into the model with the possibility of setting the value of some characteristic of the collective, a **global coordinate** s. Then it is possible to formulate the **central equation** of for the distribution $P(G|s)$, following from Bayes' theorem:

$$P(G|s) = \int d\vec{\epsilon}_G P(G|\vec{\epsilon}_G, s) \int d\vec{\epsilon}_G P(\vec{\epsilon}_G|\vec{\epsilon}_g, s) P(\vec{\epsilon}_g|s) \tag{2}$$

where $\vec{\epsilon}_g \equiv (\epsilon_{\eta_1, g_{\eta_1}}(\zeta), \epsilon_{\eta_2, g_{\eta_2}}(\zeta), \ldots)$ is the vector of the intelligences of the players with respect to their associated private utility functions, and $\vec{\epsilon}_G \equiv (\epsilon_{\eta_1,G}(\zeta), \epsilon_{\eta_2,G}(\zeta), \ldots)$ is the vector of the intelligences of the players with respect to G. The criterion $\epsilon_{\eta,g_\eta}(\zeta) = 1$ means that player η is fully rational at ζ in the sense that its move maximizes the value of its utility, given the moves of the players. A welcome feature of Wolpert's model is that consideration of points ζ at which *not all* intelligences equal 1, provides a basis for a *model-independent* formalization of bounded rationality game theory. [*ibid.*, 6n6] A point ζ at which all components of $\vec{\epsilon}_G = 1$ is a local maximum of G (a critical point of the $G(\zeta)$ surface). Let us summarize the meaning of the decomposition of $P(G|s)$ provided in the formalization:

- if we can choose s so that the third conditional probability in the integrand is peaked around vector $\vec{\epsilon}_g$ all of whose components are close to 1, then we have likely induced large (private utility function) intelligences
- if we can also choose s so that the second term is peaked about $\vec{\epsilon}_G$ equal to $\vec{\epsilon}_g$, then $\vec{\epsilon}_G$ will also be large
- finally, if the first term in the integrand is peaked about high G when $\vec{\epsilon}_G$ is large, then our choice of s will likely result in high G, as desired. [*ibid.*]

It is not necessary for our purposes to delve into the details of Wolpert's model or its applications in particular situations. The applications require an implementation of an algorithm that optimizes the signal/noise ratio reflected in the third term in the decomposition of $P(G|s)$. The signal/noise ratio signifies the uncertainty associated with a single player η who is trying to discern how its actions affect its utility $g_\eta = G$. The large number of other players and their action on G dilute η's effect on its own private utility function. Various algorithms optimize this ratio in different ways. The crucial point is that Wolpert's model of designing collectives has been tested in many different experiments. Moreover, not only have these tests "clearly validated it", but have often resulted in "performance up to orders of magnitude superior to traditional techniques from the fields of multi-agent systems and economics/mechanism design." [*ibid.*, 12] The reason for the superior performance is clearly stated by Wolpert:

> Intuitively, that superiority lies in the fact that these alternative approaches completely ignore the issue of how an agent's ability to maximize a candidate private utility will vary with changes in that private utility. This issue is especially crucial in large systems, in which each agent will face an extremely difficult signal-to-noise term in discerning the effects of its actions on its utility unless that utility is carefully crafted. [*ibid.*]

Wolpert's model of collective design provides a convincing vindication of the idea of combining the agent-based and structural aspects of action and decision-making within one coherent framework. It delivers precisely what was set as a general desideratum above, viz. a concrete model for making explicit the interactions between a single agent and the social structure within which its actions take place. The intentions of the agents are reflected in the private utility functions g_η and the structural boundary conditions in the world utility function G. The domain of applicability of Wolpert's ideas is impressive, ranging from network packet routing, the problem of controlling communication across a constellation of satellites to setting the states of the spins in a spin glass to minimize energy, the conventional bin-packing problem of computer science, and a model of human agents connected in a small-world network who have to synchronize their purchase decisions. Given the encouraging results of the mathematical testing of the model, it is likely that more interesting applications will soon emerge. It would certainly be very interesting to see, for example, whether Wolpert's model could be used for studying the selection of scientific hypotheses within a given discipline. However, the scope of this dissertation does not allow us to tackle with this question, although it is one that has a very intimate connection with the main themes of this work. The science of COllective INtelligence (COIN) is very young and it is too early to make any definite assessments about its significance for the study of collective cognition in general and about its implications for philosophical and sociological study of science in particular, including specific questions about theory choice within a scientific community. However, in as much as decision theory and game theory are arguably indispensable ingredients of a full-fledged theory of ethics, it is likely that the COIN framework will also have an important role in such a theory.

In the next chapter we turn to the discussion of the moral dimension of the formation of a scientist (or more particularly, a science-technology generalist, to use John Archibald Wheeler's apt phrase). We will discuss the significance of the interaction between the intellectual and moral dimensions of scholar's work and show how considerations of ethics (and especially ethics of ethics) have an important part to play within a unified conception of science. Furthermore, we will discuss the relevance of the historical dimension of morality – and this partly justifies the lengthy survey we have provided in the present chapter – for the transformation of the moral. Two important issues emerge within this context, viz. (i) the parallel development of valuative and theoretical thought as well as the corresponding institutional contextual conditions, and (ii) the different scales of the phylogenesis of human moral consciousness on the one hand, and the moral development of a single person, on the other. All these phenomena presuppose the distinction between the genesis and validity of ethical ideas, and the most important issue in the relationship between morals and history is to understand why socially valid norms deviate from ideal norms. It turns out that the dimensions of moral and scientific are inextricably linked and this to a large extent explains the deviations in question but also poses the greatest challenge for the intellectual and moral formation of a person in general and a scientist in particular.

Chapter 2

The Relationship between Morals and Science

2.1 Concepts of the Moral and Concepts of the Scientific

Max Weber's 'Science as a Vocation' ['*Wissenschaft als Beruf*'] [Weber (1919a)] is the *locus classicus* of the idea that science – if it is to deliver on the promise to provide the sole method of acquiring objective knowledge – must be value free.[1] The criterion of value-neutrality then became an intellectual rallying point for a large part of the European intelligentsia who vowed to the name of objectivity, fully convinced that it was not only possible but necessary to dispose of the influence of any kind of talk about values (except possibly purely epistemic ones) in scientific contexts. But however ethically neutral science and technical knowledge in general were thought to be, it cannot be denied that they critically depended on a *forma mentis* whose basic characteristic was a high capacity for abstraction that could be conducive to, not only scientific and technical knowledge, but universalist principles of ethics. In this sense the criterion of value-neutrality which often was advocated with a fully conscious, but inherently contradictory combination of modern technical knowledge and formal-universal law with anti-universalist moral ideas, can be seen as an empty deontic principle, a tautology that on the one hand does not provide any significant insight into the formal structure of experimental science and on the other hand ignores the highly relevant, albeit complex and entangled, interaction between technical rationality and moral rationality in the processes of application of scientific knowledge. Furthermore, it is neither necessary nor sufficient for nurturing a maximal capacity of technical rationality (a value in itself) that it implicitly poses as a goal. The critics of the thesis of value-neutrality (for example Hilary Putnam) have stressed the fact that all attempts to solve the problem of theory selection without the notion of epistemic values have failed. These considerations alone show that the domains of

[1]Weber was, of course, mainly concerned with sociology, but his conception was naturally intended to be deployed across other disciplines as well.

formal-universal knowledge and ideal validity have a much more intricate relationship than is often thought in the scientific context.

The task before us is to describe the relation between descriptive analysis and analysis of values. In as much as our modern scientific world conception and the general, essentially human, consciousness of the domain of validity seem to be in a fundamental conflict, a philosophical clarification of the issues that turn on this fundamental distinction has to begin with a preliminary definition of concepts. For that purpose I will briefly sketch the contours of the complementary classes of the concepts of the scientific and the moral. The emphasis is on the intension of these concepts as we understand them on a criteriological basis, i.e., we will be mainly interested in the questions of validity although genetic insights will be needed later, when we study exemplars of important moral concepts. Indeed, it is necessary to acknowledge – with respect to *both* scientific knowledge and moral positions – that the issues of genesis and validity have little in common. Both the image of nature, built upon the masses of scientific and technological knowledge gathered, and the modern conceptions of the moral have developed in the course of history. The lesson that historicism can teach us is the possibility to adopt a symmetrical attitude with respect to the status of the questions of genesis and validity within these (very different) domains. This symmetric attitude enables us to see that the validity of a theory or position (in science or in moral philosophy) cannot depend on the diachronic aspects of its genesis. Rather, it is precisely the case that the late appearance of certain scientific theories and certain moral positions is an index that they are complex and presuppose a great deal genetically, and this is seen to be a common feature of all good theories.

The point of departure for my discussion is that the terms "moral" or "morals" are used to denote strictly *normative* concepts. Therefore, when focusing on the diverse objects of philosophical inquiry such as acts, institutions, emotions, or expressions of will, we say that they are moral when they are as they *ought* to be. We have to distinguish from this, primary, meaning of "moral" the meaning "related to morals". The latter meaning is used when, for example, we investigate moral arguments that are not themselves instances of what should be but rather instruments that enable us to discover that what should be. The same goes for "moral feeling". On a more general level, if we are investigating types of behaviour, then the "morals" of these types are related to the quintessence of what they should be. The discipline that is concerned with the normative questions about human acts of will, feelings, actions and institutions, is called "practical philosophy". It is usually divided into the subdisciplines of individual ethics and political philosophy. We are here mainly concerned with individual ethics, but then in a very special context, viz., the community of scientists. It is evident that the political dimension – which is by definition the range of phenomena related to *cratics*, i.e., the study of the different uses and distributions of power – cannot be completely ignored in studies of the scientific enterprise and its ethics, in as much as the scientific enterprise itself is explicitly regulated by cratic factors, such as science policy and the decision-making procedures of both national and international funding corporations. Nevertheless, for the purposes of this dissertation, it is not necessary to delve into the details of the political aspects of ethics as my focus will be on the intellectual and moral formation of an individual scientist and the importance of an ethical self-image of scholar's vocation.

The normative dimension of practical philosophy is reflected in the view that it alone can make well-founded assertions concerning the moral nature of an action or an institution. However, a purely normative approach to ethics would not proceed very far if it did not acknowledge the practical significance of the descriptive psychological and sociological concepts that are related to people's actual ideas of of what they consider valid. In dealing with this dichotomy, we should distinguish between the notions of "ideally valid", "individually valid" and "socially valid". These terms are used in connection with different values of preferences. Individually valid or socially valid values/preferences are the subject matters of empirical moral psychology and sociology of mores.[2] As the statements of moral psychology and sociology of mores are descriptive – they fulfill the minimal criterion of value-neutrality in this sense – they are ethically neutral. This makes it possible for a philosophical analysis to discern the purely contingent features of the values of individual cultures and widespread moral feelings from the features that have ideal validity. It may well emerge from a philosophical process of justification that some such collections of values or emotions have no well-established claim to validity at all. It is thus necessary to pay attention to the *qualitative* difference between the assertions of fact and pure validity; the former are a necessary condition of a scientifically grounded ethics, but only the latter can constitute and express the hierarchy of values that is the absolute presupposition of a purely normative discipline. To see the difference, consider the following example. It is possible to distance oneself from one's own moral feelings, when it is seen that the guilt that one feels (for example) because of the incriminating talk of a superior is only a result of an unfortunate socialization and has no objective ground in itself. Moreover, such a situation provides also an example of a category mistake which is committed when the mechanism of emotion formation is exploited to someone's advantage (in this example, for the superior's advantage) without an appeal to moral arguments. These kind of category mistakes (when they are intentional) may be considered deeply immoral. From the point of view of ethics, the straightforward identification of the purely descriptive statement of moral convictions with a normative statement is an instance of extreme intuitionism. As Vittorio Hösle has observed, "[i]n European intellectual history it was the French moralists who first conceived the psychology of morals as a new discipline distinct from ethics. However, these early moralists still recognized absolute moral norms on the basis of which they criticized the mechanisms of moral pychology. This changed only with Nietzsche, who is best interpreted as a moralist without morals." [Hösle (2004), 72]

To make the central idea of the above discussion explicit, to claim that the normative dimension could be captured by descriptive theories, for example by maintaining that ethical theories could be fundamentally formulated as theories of sociology, must be con-

[2]The notion of "mores" is here adopted from Vittorio Hösle who describes the related Hegelian concepts of "morality" and "mores" as follows: "The Hegelian concepts of 'morality' (*Moralität*) and 'mores' (*Sittlichkeit*) are also used [. . .] as descriptive concepts. What I understand by the 'mores of a culture' is its ethos, that is, all its central, socially valid values, on one hand, and on the other the feeling of collective identity in particular; the latter springs from the general confidence that everyone recognizes these values. By 'morality' I understand the isolation of the individual from the concrete values of his mores, on one hand, and on the other in particular from the collective identity that springs from mores. Because these terms are used purely descriptively, we can speak of the mores of Nazi Germany or of the morality of anarchistic terrorists." [Hösle (2004), 73]

sidered a fallacy. An analogy with two different types of analysis of a given copy of a book, say Darwin's *The Origin of Species*, may be illuminating. From a purely descriptive point of view, a phenomenological analysis of the given book might result in the statement that it is nothing but a physical object of certain weight consisting of hard covers and pages made of ink, paper, parchment, or other materials and fastened together to hinge at one side, etc. A sophisticated physical analysis might provide more accurate information about the detailed properties of the materials out of which the book is made. While it is incontestable that a book can be described in these terms and that it might even be cognitively meaningful to look at the book in this way (for example for the purposes of developing more durable and environmentally friendly materials for book-binding industry), it is obvious to us that the essence of the book is not exhausted by such an analysis. Just as the dimension of meaning encoded in the physical signs goes beyond the domain of physical being, the normative dimension transcends that of 'ordinary' meaning. The Husserlian concepts of *noesis* and *noema* can be used here to clarify the issue philosophically (abstracted from their original context within philosophy of consciousness). Hösle makes the point admirably: "Noesis is the act of consciousness in which a subject grasps ('intends') something, and the noema is the object of this act, what is grasped ('intended'). Similarly, an ethical theory must be distinguished from what the theory is about. The theory itself belongs to the empirical social world and can be investigated in terms of the latter's categories. On the other hand, the theory's noema belongs to a world of pure validities, and it is with this world that ethical theory is concerned. Only the theory itself, and not its noema, can be sociologically objectivized." [*ibid.*, 74]

We have now made clear what we, in general terms, mean by the normative dimension of ethics. However, it would be erroneous to think that we have thus emancipated ethics from the domain of empirical sciences in the sense that, for example, the sociological analysis of moral behaviour becomes redundant as soon as the normative framework for the analysis of the validity of moral judgements is established. On the contrary, the descriptive and normative dimensions are seen to be complementary to each other; to try to understand the world in terms of only one of them is seriously incomplete. The relationship between ethics and sociology has, according to Hösle, "extensive analogies with Spinoza's parallelist solution to the mind-body problem. According to Spinoza, interaction between the physical and the psychic does not occur; the physical can be causally explained only the by the physical, and the psychic only by the psychic; however, it is incomplete to consider the world in terms of a single attribute." [*ibid.*, 74–75] Hösle thinks that the purely ethical and purely sociological approaches are each incomplete, notwithstanding the fact they both deal with closed systems (within different domains). It is conceivable that ethical arguments are used as instruments (this is possible because they are social facts) to promote partisan interests in a power struggle, but this does not in the least touch upon the question about their validity; we are not allowed to make judgements in the positive or in the negative whether the argument is valid. Conversely, given a theoretical analysis of the validity of arguments, we are not allowed to draw any conclusions about the effectiveness of those arguments in the social world: "for pure validities *justify* other validities, but they do not *cause* anything real." [*ibid.*]

From the complementarity of the descriptive and the normative approaches emerges the appropriate role of philosophy as the discipline which seeks to combine wisdom and science. In other words, confronted with various presentations provided by the sciences of partial areas of reality which they describe with great clarity as if they were the whole reality, philosophy tries to accommodate within the picture of reality that has been distillated from the alembic of science the notion about the reality of the psychic and the insight that the psychic element in humans historically has its center in the recognition of values that are something more than psychic or social facts. From a philosophical point of view, it can be argued that ethically neutral sociology (or science in general) commits an abstractive fallacy in neglecting the normative dimension. Behaviorism famously committed an analogous abstractive fallacy by neglecting the psychic. But whatever the causes of these historical facts are, it could still be maintained that neither ethically neutral sociology nor behaviorism has any material gaps in their explanatory apparatus. To reiterate Hösle's remark, "ethical reasons can produce no changes in the world as experienced." [*ibid.*, 76] In the case of sociology, then, it is reasonable to build theories on the basis of the assumption that only real psychic insights (social facts) into ethical reasons can act as causes and have tangible effects in the social world; and similarly in the case of behaviorism, it is reasonable to assume that every physical event related to a human body (including movements of the body) is caused by either an outer physical process or the brain state that accompanies that particular event. But apart from the material adequacy of these disciplines with respect to their explanatory capacity, they manage to provide only a partial account of reality, and thus fall short of delivering the ingredients of a world-view that satisfies the genuinely human yearning for wisdom and an intuitive sense of the world as a whole, which the archaic people seemed to possess to a far greater degree than the moderns. Though science is the most precious and wonderful achievement of modern culture, the thesis of value-neutrality and the accompanying conception of technical rationality that seem to underlie its self-image is detrimental to the development of human culture in general and science in particular. For science can be pursued successfully – and I suppose no-one contends this – only if one has an intellectual goal that guides one's work. The selection of goals is an axiological process to be sure, and to be able to work effectively, every scientist has to engage in such a process. According to the received view of value-neutrality of science (*à la* Weber) that was canonized in a particularly idiosyncratic manner in the philosophical work of the logical empiricists, foremost in the works of Carnap and Neurath, the selection process is at the outset relegated to the set of those practical questions that are external to the scientific research process *per se*. Carnap, who formulated a pragmatic distinction between questions that are *internal* to a scientific theory, or more generally, to a well-delineated scientific discourse, and those that are *external* to it, explicitly delimited questions concerning pragmatic goals outside rational scientific discourse. From the point of view of scientific ethics and general methodology of science, this was a move that had substantial reverberations for philosophy of science and philosophy in general. The distinction itself is not the crux of the problem; one could very well construct a language $\mathcal{L}_{Ev}$ (different from the object language $\mathcal{L}_{Th}$ of the theory) in which the goals of the research process can be coordinated and evaluated (comprising, say, a decision- or game-theoretical vernacular). The problem is that the values themselves would still have to be incorporated from outside of this formal framework, and thus a further language $\mathcal{L}_{Val}$ for expressing axiological content would have to be introduced. (Such languages, of course, are habitually used by moral philosophers.) From Carnap's point of view, rational discourse about values cannot be accommodated within scientific philosophy. Thus, he also abstained from developing a systematic theory of ethics.[3] As the second part of this dissertation concentrates on the details of the development of Carnap's philosophical position during the first half of the twentieth century, it is not necessary to go deeper into these issues here. Nevertheless, it can be said that Carnap is an important and interesting example of a philosopher-scientist precisely because a genetic analysis of his philosophical thought clearly points towards a tendency to gradually bring ever larger spheres of reality under the purview of rational deliberation. And this raises the question why, despite this general tendency and the growing influence of pragmatism on his thought, he never conceived the possibility of a fully elaborated ethics. The answer to this question lies, so I will argue, in the evolving conceptions of scientific self and the fundamental elements of what can be called the intellectual and moral formation of a scientist (or science-technology generalist of which Carnap constitutes an eminent example).

In the following section I will provide a brief historical survey of the emergence and development of the *forma mentis* that underlies the scientific world conception that is characteristic of our modern culture and then adduce in a little more detail the evolution and significance of the different notions of scientific self that exemplify the particular modes of thought that are properly called scientific. After this discussion we will be in a position to weave the threads of the moral and the scientific into a more organized pattern.

2.2 The Emergence and Development of the *forma mentis* of Universalism

The notion of universalism is so important for the development of Western culture, both in terms of scientific knowledge and moral insight, that it would be impossible to form a balanced picture of the relationship between the concepts of the scientific and the moral without an analysis of the genealogy of this notion. It is therefore necessary to present, at least in rough outline, the most important aspects of the historical developments that have led to the modern conception of science and the ideal of ethical universalism. The universalism of the formal-universal laws of science has its roots already in the antiquity, but it was first articulated as a general philosophical principle in the Enlightenment. It is no coincidence that in the same era the notion was also deployed in the domain of validity and thus the Western world witnessed the emergence of the first great systems of moral universalism such as Spinoza's *Ethics*.

[3]Of course, this does not imply that Carnap's attitude to life and his character could be evaluated on the basis of this fact. Carnap had a refined moral character and he possessed moral virtues that many of his colleagues lacked. However, from a philosophical point of view, because of his idiosyncratic attitude towards moral philosophy he did not possess *ethical virtue*, i.e., an ability to defend the virtues he exemplified in his practical life in the framework of an ethical theory.

I will begin with a sketch of the history of the conception of formal and scientific knowledge and its presuppositions as they were conceived by the early Enlightenment philosophers and scientists. As a word of motivation for readers who have reservations about the utility of a genetic approach to philosophical problems I would like to point out that, in general, an examination of the historical proposals for explicating a given concept, which constitute a space of possibilities for determining the intension of the concept (in this case the concept of universalism), opens up conceptual possibilities relevant for contemporary explicative tasks within the same domain. Such historically informed inquiries are next of kin in spirit to Ernst Mach's historico-critical studies in the history of science. The utility of historically informed studies of the conceptual foundations of a science were clearly underlined in Mach's *Die Mechanik in ihrer Entwicklung — historisch-kritisch dargestellt*:

> the historical investigation of the development of a science is most needful, less the principles treasured up in it become a system of half-understood precepts, or worse, a system of *prejudices*. Historical investigation not only promotes the understanding of that which now is but also brings new possibilities before us by showing that which exists to be in great measure *conventional* and *accidental*. From the higher point of view at which different paths of thought converge we may look down upon us with freer vision and discover routes before unknown.[4] [Mach (1963) [1933], 278–279]

This approach to enquiry encompasses both technical sophistication and acute historical awareness. Thus, from the point of view of a historian of science, the historical and systematic aspects of enquiry are inextricably related. The Machian historico-critical method performs a role similar to the practice of polishing the optical surfaces in a *Schmidt-Cassegrain* telescope; when the polishing is done with care, the telescope provides an extensive and nearly flawless visual field for observation. The analogy is singularly exact. The quality of the performance of any idea depends in equal measure on the refining analyses of the historians and philosophers that use an idea for different purposes, as the quality of the telescope depends on the conscientiousness of the mirror-polishers. Consider the following sketch as just such a polishing of an optical surface.

The mathematical and methodological background of the advance of sciences in the Enlightenment

Within history of ideas, the history of mathematics has a pride of place. Studying the evolution of mathematical ideas and assessing their relevance for particular (non-mathematical) conceptual issues under investigation often provide a fruitful point of departure for enquiry. The mathematically informed mind is usually able to disclose non-articulated semantical underpinnings and possible residual meanings of given notions that could otherwise remain unnoticed. Alfred North Whitehead has expressed the

[4] Italics in the original.

importance of the history of mathematics for the history of ideas eloquently in the form of a wonderful simile (This is a passage from his Lowell Lectures of 1925):

> Even now there is a wavering grasp of the true position of mathematics as an element in the history of thought. I will not go so far as to say that to construct a history of thought without a profound study of the mathematical ideas of successive epochs is like omitting Hamlet from the play which is named after him. That would be claiming too much. But it is certainly analogous to cutting out the part of Ophelia. This simile is singularly exact. For Ophelia is quite essential to the play, she is very charming — and a little mad. Let us grant that the pursuit of mathematics is a divine madness of the human spirit, a refuge from the goading urgency of contingent happenings. [Whitehead (1946), 26-27]

Thus, following Whitehead, we can convince ourselves that it is important to take note of the insights of a mathematically informed history of ideas. There are many reasons for this, but I mention only the most important one, which perhaps is also the most obvious. The formal language of mathematics provides an incredibly efficient framework for formulating concepts in a way that allows an exact analysis of their intension. In other words, the importance of mathematics is tantamount to its capacity to act as a medium for explicating concepts, both in terms of their content as well as their inferential role. The process of mathematical theory construction – along with the invention of new mathematical concepts – is a paradigmatic exemplar within the class of strategic activities associated with meaning constitution. Its strict criteria for definability, requirement of consistency and pragmatic restrictions upon the meaning-conferring acts with respect to novel concepts gives mathematics a unique place among the social practises of man. Thus, history of mathematics provides a general background framework against which the development of ideas in general can be observed, and their logical and semantical properties analyzed.

The importance of mathematically informed history of ideas is of course pronounced in the case of Enlightenment. It would be very difficult to make sense of the radical and vast conceptual transformations of that era without taking into account the exuberant progress of the mathematical sciences in the seventeenth and eighteenth centuries. According to d'Alembert, one of the most distinguished and brilliant harbingers of Enlightenment, the eighteenth century is most aptly viewed as "the century of philosophy *par excellence*".[5] [d'Alembert (1965)[1759], 9] In fact, this statement conveys much more than a mere rhetorical point. It was a realistic assessment of the state of science and of the opportunities ahead. d'Alembert's excitement is palpable in the nearly ecstatic observations about the state of the sciences and the extent of knowledge exemplified by eighteenth century intellectual culture at large:

> [...] If one considers without bias the present state of our knowledge, one cannot deny that philosophy among us has shown progress. Natural science

[5]"Notre siécle s'est donc appelé par excellence *le sieclé de la philosophie* [...]"

> from day to day accumulates new riches. Geometry, by extending its limits, has borne its torch into the regions of physical science which lay nearest at hand. The true system of the world has been recognized, developed, and perfected ... In short, from the Earth to Saturn, from the history of the heavens to that of insects, natural philosophy has been revolutionized; and nearly all other fields of knowledge have assumed new forms ...[6] [*ibid.*, 9–10]

Whereas d'Alembert is justifiably the main representative of the Cosmopolitan Republic of Letters, the progress, which by any measure had been considerable, was mainly due to the labours of two great men who had lived and made their lasting contributions to science in the previous century. Without Descartes and Newton the victories which d'Alembert so exuberantly announced would have been inconceivable.

What, in particular, can be said about Descartes's and Newton's contributions with respect to the notion of universalism? In the first place, they were the most prominent methodologists of their time and virtually determined the foundations of scientific methodology for centuries to come. Their contributions constitute a significant part of the modern conception of the universalism of formal empirical science. Nevertheless, in spite of the shared universal tendency, they differed in their methodological thinking with regard to many substantial issues. Cassirer points out a significant distinction between the methodological approaches of Newton and Descartes that underlies the differences in the many characteristic ways of tackling philosophical and scientific problems in the Enlightenment. Newton's 'Rules of Philosophizing' (as explicated in *Principia* and *Optics*) and Descartes' *Discours de la Méthode* (and equally, his *Regulæ ad directionem ingenii*) had an enormous influence on the methodological orientation of individual thinkers and scientist of that era. Firstly, we may consider the peculiarities of Newton's method, which only recently have received the careful attention they really deserve. Newton, as is well known, used a certain variant of the method of analysis in carrying out his investigations in the natural sciences. Until quite recently, the constituents of this method remained to a large extent uncharted. Already Cassirer notes how idiosyncratic the notion of "phenomenon" is in Newton's philosophical vocabulary: for Newton, phenomena are the data of experience. [Cassirer (1951), 7] But important as the observation is, it is only a starting point for a more adequate description of the notion.[7] The important insight with respect to the Newtonian notion of "phenomenon" is that the data of experience must be structured in a determinate manner, i.e., (1) they must be acquired by means of

[6]"Si on examine sans prévention l'état actuel de nos connoissances, on ne peus disconvenir des progrés de la philosophie parmi nous. La science de la natur acquiert de jour nouvelles richesses; la géométrie en reculant ses limites, a porté son flambeau dans les parties de la physique qui se trouvoient les plus prés d'elle; le vrai syst ème du monde a été connu, développé et perfectionné; [la même sagacit é qui s'êtoit assujetti les mouvements des corps cêleste, s'est portée sur les corps qui nous environment; en appliquant la gêomêtrie à l'étude des ces corps, ou en essayant de l'y appliquer, on a su apercevoir et fixer les avantages et les abus de cet emploi;] en un mot depuis la terre jusqu' à Saturne, depuis l'histoire des cieux jusqu'à celle des insectes, la physique a changé de face. Avec elle presque toutes les autres sciences ont pris une nouvelle forme, [et elles le devoient en effect. Quelques réflexions vont nous en convaincre.]"

[7]The most important references in this respect are [Harper (2002)], [Hintikka (1992)], [Smith (2002)] and [Stein (1991)].

a controlled experiment, and (2) they have to be taken as a whole, or more precisely, as a structured aggregate governed by functional relationships.[8]

It is well known that Newton was not explicit about his methodology in the *Principia*. Although there is a famous passage right at the beginning of the book that eloquently describes the goal of the treatise in terms of laying the basis for a completely new form of natural philosophy, the actual development and elaboration of these ideas remains curiously scant in the main body of the book itself.[9] The goal of Newton's investigations, the end point of analysis, is constituted by the discovery and formulation of the principles of the particular science under consideration. The reversed order of the investigation with respect to the logical order of things derives from the epistemologically unique *point de vue* of ours: if the latter (principles of science) are first according to nature (πρότερον τῃ φύσει), then the former (phenomena) must always be first to us (πρότερον πρός ἡμᾶς). [Cassirer (1956), 7]

The covert methodological principles rehearsed by Newton in the physical sciences which he refrained from making explicit, have been, as mentioned above, a subject of heated debate during the last twenty years or so. Another, equally contested subject has been the relation of Newton's discoveries in the differential and integral calculus to modern concepts of mathematical analysis. Both cases demonstrate that it is recommendable to adhere to the golden rule of history of science: always study the methodology of a scientist in the first place directly from his scientific works, and not from his methodological pronouncements. This rule is indispensable when one studies works of scientists whose contributions have radically changed the intellectual scene of their respective fields of specialty. As far as Newton's mathematical discoveries are concerned, we can acquire an enhanced view of the whole process of conceptual growth and change during the Enlightenment and its far-reaching consequences, if we dwell in some more detail on Newton's fundamental ideas about the differential and integral calculus. In the first place, we will see how Newton is surprisingly close to Leibniz regarding his opinions about the philosophical import of the radically new techniques of calculus. This fact is related to

[8]One cannot but admire the hermeneutic abilities of the most prominent historians of science in disclosing the depths of certain important 'seed conceptions' of scientists that have been obscurely communicated in the original publications or writings. Such pieces of detective work are a constant source of inspiration. In this respect the insight that essential to Newton's method of analysis is that the data of experience must not be conceived as distinct, atomic instances of the observables, but as a structured whole, seems to me to be not only exceptionally creative, but also materially correct. (In *Principia*, the heuristic role of Kepler's laws in this structuring process were crucial, as Newton himself made clear.) Cf. especially [Smith (2002)].

[9]The famous prefatory remarks read: "And therefore our present work sets forth mathematical principles of natural philosophy. For the whole difficulty of philosophy seems to be to find the forces of nature from the phenomena of motions and then to demonstrate the other phenomena from these forces. It is to these ends that the general propositions in Books 1 and 2 are directed, while in Book 3 our explanation of the system of the universe illustrates these propositions . . . If only we could derive the other phenomena of nature from mechanical principles by the same kind of reasoning! For many things lead me to have a suspicion that all phenomena may depend on certain forces by which the particles of bodies, by causes yet unknown, either are impelled toward one another and cohere in regular figures, or are repelled from one another and recede. Since these forces are unknown, philosophers have hitherto made trial of nature in vain. But I hope that the principles set down here will shed some light on either this mode of philosophizing or some truer one." [Newton (1999), 382f.]

the gradual deterioration of purely speculative philosophical elements in the exact sciences in the 18th century, and hence to the increasing importance of the purely formal universalism within the methodology of science. Indeed, as both eschew the notion of a 'real differential' and explicitly reject a metaphysical interpretation of the techniques of analysis, mathematical physics is relieved of the redundant metaphysical packaging which still characterizes the work of Descartes. In the second place, the formulation of the basic concepts of calculus demonstrate clearly what kind of standards of rigor the mathematicians of Newton's time advocated. This is important both from the point of view of seeing the development of mathematics as an important background framework for cultural evolution at large, and from the point of view of the internal development of mathematics itself and the processes of explication rehearsed under its purview.

The differential calculus emerged in a gradual process of evolution in which Newton and Leibniz played the chief roles. Both grounded their work on the achievements of such prominent mathematicians of previous generations as Kepler, Cavalieri, Descartes and Fermat. Proclaiming the merit of the actual discovery of the calculus to both Newton and Leibniz is not ungrounded, however, since they first reduced the solution of the cornucopia of problems their predecessors tackled by methods of analysis of infinitesimals, to a systematic use of the two mutually inverse operations of differentiation and integration. As far as matters of priority are concerned, in the sense of printed publication the honour belongs to Leibniz, who is known to have given a comprehensive account of the fundamental problems and methods of the new calculus in a series of papers published in the journal *Acta Eruditorum* between 1682–1686. Nevertheless, Newton had come up with his formulation of the fundamental ideas earlier in the unpublished papers (1) *De analysi aequationes numero terminorum infinitas* in 1665; (2) *Methodus fluxionum et serierum infinitarum*, after the *De analysi per aequationes*, but before 1671; and (3) *De quadratura curvarum*, the main text in 1665–1666. These works were not published until (1) in 1711; (2) after Newton's death in 1736; and (3) in 1704 as an appendix to *Optics*.

It is expedient to make clear the difference between their approaches to settle the first issue concerning their different opinions about the philosophical implications of calculus. Alekseĭ Nikolaevich Krylov, a Russian naval engineer and applied mathematician, and the first author to translate Newton's *Principia* into Russian, has described this difference with characteristic clarity:

> Netwon discovered and gave the bases for the calculus of infinitesimals by starting from mechanical and geometric concepts. In his arguments he always applied geometric notions and was absolutely rigorous in them and absolutely precise in language and expressions. Therefore, he first establishes the concept of the limit of a variable quantity, the concept used today, and all his teaching about "fluxions", or "derivatives" in modern terminology, is based on finding the limits of ratios of two infinitesimal quantities that are in a definite mutual dependence and that vary jointly. Posing as the basic problem of integral calculus the determination of a "fluent" from its given "fluxion", that is, of a primitive function from its given derivative, he con-

stantly employs geometric notions, and his tract itself is called *De quadratura curvarum*.

> Leibniz proceeded in a different way. He introduced the new term "infinitesimal" instead of an increment of a variable or a function of it that vanishes in the limit. He did not give this concept a precise and rigorous mathematical definition, and in some of his explanations he seemingly did not distinguish between the mathematical concepts of "infinitely small" and "very small", nor between "infinitely large" and "very large", likening, for example, the one to a speck of dust and the other to the Earth. Moreover, he connected the concept of an infinitesimal with the philosophical concepts of "finite or infinite divisibility of matter", of an "indivisible atom", of a "monad", and so on, concepts very far from pure mathematics, which has to do not with the quantities themselves but with numbers serving as a measure of them. [Krylov (1933), 16]

Although this description captures some of the essential differences between Newton and Leibniz, it tends to make the distinction between them too sharp. In the first place, Newton and Leibniz very much share the fundamental presuppositions of the then current mathematics, and their calculi are strongly embedded in the culture of their own times. Two points suffice to illustrate this. (i) Neither Newton's nor Leibniz's calculi are about "functions" which should be evident because the notion of function emerged only later, particularly in the work of Euler and Cauchy. Instead of "functions" Newton and Leibniz talk about "quantities", and the references they make to these quantities, their rates of change, their differences, etc., are always accompanied with specific geometric constructions or entities (typically curves). (ii) The notion of continuum that we habitually associate with calculus, the continuum of the real numbers, is not employed by Newton and Leibniz. Rather, they rely on a geometrical or kinematical notion of a continuum. Referring to an intuitive geometrical or kinematical continuum, they are able to develop their limit procedures needed in analysis. From a philosophical point of view, one of the most important foundational questions of mathematics in the seventeenth and eighteenth centuries was the question concerning the referential content of mathematical symbolism. The question "Do differentials exist?", that is commonly taken to be the demarcation criterion between the conceptions of analysis of Newton and Leibniz, turns out to be, on closer analysis, a question to which both Newton and Leibniz reply in the negative. For, as Niccolò Guicciardini has remarked: "they both stated that *(a) actual infinitesimals do not exist; they are useful fictions employed to abbreviate proofs, (b) infinitesimals should be defined rather as varying quantities in a state of approaching zero, (c) infinitesimals can be completely avoided by limit-based proofs, which constitute the rigorous formulation of calculus, (d) limit-based proofs are a direct version of and are thus equivalent to the indirect,* ad absurdum, *Archimedean method of exhaustion*." [Guicciardini (2003) [Jahnke (2003), 97]] What distinguishes Leibniz's view from Newton's is the *philosophical justification* that Leibniz provides for the application of infinitesimal quantities as "well-founded fictions".[10] Especially in his later writings Leibniz argued that differentials are well-founded, since they are symbolic abbreviations for limit-procedures. On one occasion he wrote:

[10] Drawing essentially on his general philosophical conception of *phenomena bene fundata*.

> In fact, instead of the infinite or the infinitely small, one can take magnitudes that are so large or so small that the error will be less than the given error, so that one differs from the style of Archimedes only in the expressions, which are, in our method, more direct and more apt for the art of discovery. [Leibniz (1701), 350][11]

These considerations express concisely the dialectic between the conceptions of two of the most remarkable natural philosophers of the seventeenth century. The important point is not the dispute that raged between these two men about the priority of the invention of the differential and integral calculus, but the simultaneous assessment of two qualitatively different *proposals* for the basic system of differential calculus. The central contributions of both men comprise three aspects of mathematical discovery: problem-reduction, the calculation of areas by inversion of the process for calculating tangents, and the creation of algorithms. Observing the issue from our modern point of view, it is now clear that aspects of both men's contributions have been axiomatic for the development of mathematics. The invention of infinitesimal calculus constituted the first great technical revolution in the course of the development of formal-universal laws; it was the language in which the laws of Newton in particular and the laws of classical physics in general were most naturally expressed.

Apart from the strictly technical work relating to pure mathematics, the most spirited methodological debate concerning the justifiable procedures of knowledge acquisition in science ignited between the ideas of Descartes and Newton. Newton's method we have already briefly sketched above. It is necessary to supplement this description with a brief account of Descartes' methodological principles.

Descartes, as is well known, dedicated a considerable part of his early career to studying methodological questions. The first attempt in this vein is the famous *Regulæ ad directionem ingenii* which Descartes never finished. It had apparently been written in 1628, though it was published only posthumously. It comprises twenty-one so-called rules, but it is probable that a number of other rules were also premeditated as can be deduced from other textual evidence. The essentials of the *Regulæ* consist of principles that govern the use of the two most important mental operations that men have recourse to in their cogitations, i.e., *deduction* and *intuition*. The method consists in rules for employing aright these two mental operations, and it is said above all to consist in order.

In the fifth of the rules of the *Regulæ* Descartes gives a summary of his method: "Method consists wholly in the ordering and disposing (literally, in the order and disposition) of those objects to which the attention of the mind must be directed if we are to discover any truth. We shall observe this method exactly if we reduce involved and obscure propositions step by step to those which are simpler, and if we then start with the intuitive apprehension of the simplest propositions and try by retracing our path through the same steps to ascend to the knowledge of all the others." [*Regulæ ad directionem ingenii*, §5; AT X, 379] Although it may not be immediately clear, a little reflection shows that Descartes

[11] Translated in [*ibid.*, 98].

is in fact talking here about the method of analysis. The desiderata of this method are related to the two precepts that Descartes introduced in his *Discours de la Méthode*. The first part of the method is that we should reduce involved and obscure propositions step by step to those which are simpler. And this injunction is generally said to correspond to the second precept of the *Discours*: "The second (precept) was to divide up each of the difficulties which I was to examine into as many parts as possible and as seemed requisite." [*Discours de la Méthode*, 2; AT VI, 18] This method has become to be known by the names of *analysis* or *resolution*. It is a contested issue whether Descartes always used the term "analysis" in precisely the same sense, but the sense conferred to it here consists in breaking down, as it were, the multiple data of knowledge into their simplest elements or element. Needless to say, Descartes was very much influenced by the ideas he had encountered in mathematics in scrutinizing the concept of analysis. Indeed, the work that Descartes did in analytic geometry was in large part inspired by the desire to re-evaluate and refine the notion of exactness in mathematics through acquiring a deeper understanding of what was involved in the classical conception of analysis that the ancient Greek geometers had recourse to in their work. Especially important was the relationship between the idea of analysis and the notion of construction. But according to Descartes, Euclidean geometry, for example, had a serious drawback, namely, that the axioms and first principles are not 'justified'. That is to say, the geometer does not show how his first principles are reached. The upshot of the methodological investigations of Descartes is that using the method of analysis or resolution, one can 'justify' the first principles of science. This is accomplished by making it clear in a systematic manner (this is precisely the underlying purpose of the method) *how* they are reached and *why* they are asserted. This makes it clear that analysis is in a sense a *logic of discovery*. I will touch upon this particular issue in the next chapter.

The second part of the method that is summarized in the fifth Rule in *Discours de la Méthode* insists that we should "start with the intuitive apprehension of the simplest propositions and try by retracing our path through the same steps to ascend to the knowledge of all the others." This is essentially an account of the method of *synthesis* or *composition*. We thus see how the two 'directions' of inference are related. Analysis can most naturally be seen as a method for discovery, whereas synthesis is best suited for demonstrating what is already known. The dialectic of these two methods constitutes actually one method the proper functioning of which essentially depends on the ordered and systematic application of both the 'upstream' and 'downstream' modes of inference represented by analysis and synthesis. As we have seen, Newton's argumentation in both mathematics and physics exemplifies the critical importance of these methods. It is therefore apt to conclude that these two men, Descartes and Newton, were the two great analysts of the seventeenth century.

The ideological roots of Enlightenment

> I. Say first, of God above, or man below,
> What can we reason, but from what we know?

> Of man what see we, but his station here, From which to reason, or to which refer?
> Through worlds unnumber'd though the God be known,
> 'Tis ours to trace him only in our own.
> He, who through vast immensity can pierce,
> See worlds on worlds compose one universe,
> Observe how system into system runs,
> What other planets circle other suns,
> What varied being peoples ev'ry star,
> May tell why Heav'n has made us as we are.
> But of this frame the bearings, and the ties,
> The strong connections, nice dependencies,
> Gradations just, has thy pervading soul
> Look'd through? or can a part contain the whole?
> — Alexander Pope, *An Essay on Man*, Epistle I

Enlightenment historiography typically draws on rhetorical phrases like "an illustrious age of reason" when describing the characteristics of Enlightenment thought. This is admittedly a succinct way of stressing the importance of the methodology of natural philosophy (as science was then called) in bringing forth a heightened capacity of technical rationality which made possible all the progress in the technical infrastructure as well as in the administrative and economic systems of Western societies. Furthermore, these accounts regard the so-called 'Scientific Revolution' as a significant catalyst in the overall ideological development during this era, accelerating the 'plate tectonics' of the various strata of philosophical thought *in statu nascendi*. Although this portrait of Enlightenment certainly captures an important aspect of its essential character, it is only a partial truth. The situation was more complicated than this standard picture makes it seem. The complexities derive from at least three different sources: (i) the 'Scientific Revolution'[12] was only one of the background influences of the intellectual transformation that largely took place during the half-century between 1670 and 1720 (Indeed, the 'Scientific Revolution', having taken place, roughly, between Galileo and Newton, was already over by that time.); (ii) the notion of revolution that came to play such a prominent role in the Enlightenment thought had already been associated with an earlier major intellectual influence, Cartesianism; it was not, thus, as is usually argued in contemporary Enlightenment historiography, a product of the political surges culminating in the French Revolution in 1789; (iii) there were, right from the beginning, two enlightenments, not one; the dialectic between the conservative and the radical forms of enlightenment comprises one of the most important chapters in the history of seventeenth- and eighteenth-century thought. Let us elaborate on these aspects in a little more detail.

¶In the first place, it is contested among a number of historians of science that there ever was a sharply defined and unambiguously identifiable 'Scientific Revolution' encompassing a radical change in the theoretical tools, experimental techniques and means

[12] An anachronistic expression, because the discipline under the purview of which the radical intellectual shift took place was not known as "science", but rather as "natural philosophy".

of communication between colleagues within natural philosophy. What these sceptical evaluations of the status of the 'revolution' have in common is the tendency to point out scholars, mostly from the Medieval era, that already, not only anticipated, but put to concrete use the methodological tools and techniques that according to the common view were only developed during the period from Galileo to Newton. Although there certainly is a sense in which it is justifiable to talk about some exceptionally foresighted individuals who devised tools the subsequent development of which led to the widely adopted forms of mathematical presentation and techniques of calculation (Nicole Oresme's (1320–1382) use of graphs to illustrate the dependencies between variable magnitudes is an example), experimental procedures, modes of presentation, etc., it is a well-established fact that the work of such scientists as Galileo, Kepler and Newton would have been inconceivable without the radical change in the two fundamental disciplines that underpin the 'revolution', viz., geometry, and the theory of free fall and compounded motion. On the one hand, the redefinition of the notion of exactness in mathematics is largely due to Descartes' transformation of the early modern concept of geometrical construction. This work underlies the development of analysis in the works of Newton and Leibniz, to which we have had occasion to refer above. The significance of mathematical analysis for the study of physics and its rapid development is, of course, beyond question. On the other hand, the investigations of free fall and compounded motion by Descartes, Beeckman and Galileo constitute the background for the subsequent development of mechanics. It is mostly because of their pioneering efforts in "exploring the limits of preclassical mechanics" [Damerow et al. (2004)] that the theoretical discoveries of Huyghens and Newton were made possible in the seventeenth century. There are, then, strong systematic reasons for maintaining that the Scientific Revolution was a phenomenon confined to a specific, identifiable historical period. Descartes made the necessary preparations and Galileo implemented the new ideas stemming from the study of motion, presenting the first coherent formulation of the laws of free fall and compounded motion. In addition to the systematic reasons, there is historical evidence that the revolution was, indeed, a fact. Modern historiography would be at pains to explain the systematic references to a "Scientific Revolution" in the literature of the early eighteenth century, if there was no factual transformation of the modes of thought in the seventeenth-century natural philosophy. As Jonathan Israel has put it: "[T]he recent trend among the historians of science to question whether there really was a 'Scientific Revolution' of discoveries, new procedures, and instruments which fundamentally changed the substance of scientific debate in the seventeenth century leaves untouched the vast influence of the early eighteenth-century perception of 'Scientific Revolution'. Indeed, it would seem to strengthen the argument [that the modern concept of 'revolution' is specifically a product of the Early Enlightenment] that it is precisely in the 'displacement of the conceptual network through which scientists view the world' by an essentially new paradigm, a change in categories and ideas, a philosophical transformation in other words, that one finds the really significant difference between what is pre-modern and what is 'modern'." [Israel (2006), 6]

¶In the second place, the notion of "revolution", in the sense of a principled, general discarding of authority and traditional premisses, was largely instigated in Europe by the advent of the mechanistic world-view of Descartes. Cartesianism, which triumphed widely in the later seventeenth century, marked a great shift in basic concepts (in philosophy and mathematics alike, as we have seen), thereby changing western civilization profoundly, and, as noted above, among innumerable other changes, transforming the meaning of the word "revolution" itself. Descartes advocated a reformation of the whole system of knowledge, embarking on a transformation that was aptly called by Turgot in 1748, "une révolution totale." This description was wholly justified in as much as Descartes' notion of "revolution" means, as Jonathan Israel has put it, "not just linear, fundamental, and irreversible change, and not just auto-emancipation from the intellectual and cultural shackles of the past, but also, as Turgot's remark indicates, something that changes everything." [*ibid.*, 5]

¶In the third place, it is essential to emphasize the importance of the intellectual import of the radical enlightenment as compared to the contributions of the conservative one. Although the dichotomy between radical and conservative Enlightenment – or moderate mainstream Enlightenment, as Jonathan Israel likes to call it when dwelling upon this contrast – has been contested by some scholars, there are compelling reasons to accept it as a correct description of the intellectual situation in Europe between 1670 and 1720. To see this, it is sufficient to point out its strategic significance in historical studies that stress the importance of *intellectual* impulses as a relevant category of the explanatory apparatus along with cultural, social and material factors. What, then, is the difference between "radical Enlightenment" and "moderate mainstream Enlightenment"? Jonathan Israel's seminal monographs, *Radical Enlightenmet* [Israel (2001)] and *Enlightenment Contested* [Israel (2006)], which provide a remarkably comprehensive analysis of the various dimensions of Enlightenment and present a sustained argument for the relevance of the "two enlightenments" for general history of ideas, are not alone responsible for the currency of the line of Enlightenment historiography which makes the contrast between radical and conservative enlightenments its central explanatory and heuristic template. Neither does the phrase "radical Enlightenment" derive from Israel. The significance of the so-called 'radical' wing of Enlightenment had already been emphasized by Margaret C. Jacob in the early 1980s, albeit from a slightly different perspective. In the light of historical evidence it is incontestable that there were *several* strands or varieties of "radical Enlightenment" and that most of them were intellectually secluded from one another in terms of direct communication or cultural influence. Nevertheless, they had quite a few characteristics in common. In the most general terms, one may associate these varieties of radical Enlightenment with a critical attitude towards the thought of French Enlightenment *philosophes* – particularly that of Voltaire and d'Alembert:

> The radicals were intellectual dissenters, men, and possibly a very few women, often with a Protestant refugee background, who could not share the willingness of the major *philosophes* like Voltaire and d'Alembert, or liberal churchmen like the Newtonians in England, to put their faith in enlightened monarchy. They sought, therefore, through a variety of methods, propaganda as well as intrigue, to establish a republican ideal, if not always a republican reality, worthy of European-wide imitation. Predictably they, like the moderates, were the intellectual heirs of the mid-century English Revolution, only unlike the moderates they sided more with the radical sectaries, that is, with

> the losers rather than the winners of that first major European revolution. In Continental Europe many intellectual legacies were at work in the later seventeenth century – Cartesian, Spinozist, Grotian to be sure – but the great revolution across the Channel must also be acknowledged for the inspiration that the discontent and the angry drew from it. [Jacob (1981/2003), ix]

The division of the advocates of Enlightenment values into different camps resulted from different basic attitudes towards the application of the criteria of rationality. Few sought to apply the new criteria to everything. Indeed, as Jonathan Israel has noted: "[j]ust as Descartes with his two-substance dualism created a reserved area of spirits, angels, demons and miracles, and Boyle and Locke with their emphatic empiricism similarly ring-fenced miracles, spirits and the core Christian 'mysteries', so the intellectual elites of Europe mostly sought one or another intellectual expedient for having it both ways – that is reconciling the new mechanistic criteria of rationality not just with religion and theological doctrines but also with social norms and notions of education, society and politics based on custom, usage, and existing laws as well as social-hierarchical principles." [Israel (2006), 10] This was, then, the fundamental demarcation criterion between the two variants of enlightenment: whether reason alone reigns supreme in human life or whether philosophy's scope must be limited and reason reconciled with faith and tradition.

> From the outset then, in the late seventeenth century, there were always two enlightenments. Neither the historian nor the philosopher is likely to get very far with discussing 'modernity' unless he or she starts by differentiating Radical Enlightenment from conservative [...] moderate mainstream Enlightenment. For the difference between reason alone and reason combined with faith and tradition was a ubiquitous and absolute difference. Philosophically, 'modernity' conceived as an abstract package of basic values – toleration, personal freedom, democracy, equality racial and sexual, freedom of expression, sexual emancipation, and the universal right to knowledge and 'enlightenment' – derives [...] from just one of these two, namely, the Radical Enlightenment; historically, however, 'modernity' is the richly nuanced brew which arose as a result of the ongoing conflict not just between these two enlightenments but also (or still more) between both enlightenments, on the one hand, and, on the other, the successive counter-enlightenments, beginning with Bossuet and culminating in Postmodernism, rejecting all these principles and seeking to overthrow both streams of Enlightenment. Rousseau, initially in the late 1740s and early 1750s an ally of Diderot and radical *philosophe*, subsequently, in the 1760s, rebelled against both branches of Enlightenment, becoming the moral 'prophet' as it were of one form of Counter-Enlightenment.
>
> Of the two enlightenments, the moderate mainstream was without doubt overwhelmingly dominant in terms of support, official approval, and prestige practically everywhere except for several decades in France from the 1740s onwards. Nevertheless, in a deeper sense, and in the long run, it proved

> to be much the less important of the two enlightenments. For it was always fatally hampered by its Achilles heel, namely that all its philosophical recipes for blending theological and traditional categories with the new critical-mathematical rationality proved flawed in practice, not to say highly problematic and shot through with contradiction. Cartesian dualism, Lockean empiricism, Leibnizian monads, Malebranche's occasionalism, Bishop Huet's fideism, the London Boyle Lectures, Newtonian physico-theology, Thomasian eclecticism, German and Swedish Wolffianism, all the methodologies of compromise presented insuperable disjunctions and difficulties, rendering the whole philosophico-scientific-scholarly arena after 1650 exceedingly fraught and unstable. [*ibid.*, 11]

Surely, the division between "radical" and "conservative" Enlightenment (as pepetuated by Israel and Jacob) is an idealization and therefore considerably simplifies the complex and nested hierarchy of interrelated facts and impulses that constituted the cultural environment of Enlightenment, but then, this qualification could be presented with respect to most heuristic divisions in the history of ideas. If our goal was purely historiographical and confined solely to the history of Enlightenment thought, we would be obliged to delve more deeply into the details of historical evidence and definite presuppositions of the interpretative framework provided by Israel and Jacob. The intention here, however, is not to explicate the content and complexity of the notion of Enlightenment *in toto*, but rather to bring forth selectively such heuristic insights into the notion of Enlightenment and its history that amend and supplement the somewhat one-sided view of Enlightenment that the conventional Enlightenment historiography presents us with. The division of Israel and Jacob is not to be accepted dogmatically, but rather as a working hypothesis which sheds some light upon the important differences between the basic philosophical presuppositions and stances of the numerous thinkers that are usually collected under the umbrella term "Enlightenment philosophers". Moreover, as far as Spinoza and spinozism are concerned, the division of Israel and Jacob seems to be highly relevant in the context of the debate on tolerance that involved Locke and Spinoza in the late 1680s. We shall have occasion to return to this debate in section **2.4** where we study the contours of the development of moral universalism.

In as much as the purpose of this historical discussion is to dwell especially upon the genealogy of certain important philosophical notions – in particular, the notion of universalism in scientific and ethical thought, and epistemic virtues such as objectivity, truth-to-nature and trained judgement on the one hand, and moral virtues such as forbearance, permissiveness, mutual respect and tolerance on the other – the emphasis upon intellectual impulses cannot possibly be regarded as an epistemic vice. Neither is it a retreat to an "old-fashioned" variety of historiography.[13] A consensus among historians seems to be that the explanatory basis of history must primarily consist in economic and social

[13] Although some references mentioned in the text date back to the early 1900s, it is neither substantially correct nor moral to criticize the intellectually oriented approach to history as "old-fashioned". Grounding these judgements solely on chronological characterizations alluding to the year of publication of a literary source miss the crucial distinction between genesis and validity. That certain research strategies are able to maintain their currency for long periods of time is an expression of the fact that questions of validity cannot

structures, or other well-defined social and cultural factors the dynamics of which can be made explicit. Investigations that focus on the intellectual dimension of a significant historical change are often regarded with skepticism or deemed "old-fashioned" right out of hand. This is an ideological presupposition that does not bear much methodological weight. Indeed, the trend of playing down the importance of ideas in the course of historical change has resulted in a desultory state of historiography, as Israel has correctly stressed: "If, moreover, in recent decades most historians of both Enlightenment and the French Revolution have repudiated interpretations emphasizing the role of ideas, claiming the revolutionary movements were primarily social and cultural phenomena best understood by focusing on social relations and material factors, there remain formidable unresolved difficulties with this conception. No one has been able to specify what the allegedly profound social changes which lay behind the Enlightenment and Revolution actually were or even how shifts in social structure, given their reality, could broadly and spontaneously translate into popularly driven 'universal revolution' designed to transform the core principles upon which society and politics rest." [Israel (2006), x] However, there are not just negative reasons for abandoning the conventional presuppositions of historiography. Much more is at stake. The "controversialist" approach to intellectual history advocated by Israel, fusing the advantages and shedding the disadvantages of both the conventional social, cultural and political history, and intellectual history, will – so Israel claims – amount to a unique cultural phenomenon of considerable importance to a wide class of scholars and lay readers, including not only historians, but also philosophers, social theorists and political analysts. The task is not an easy one. As Israel puts it: "To integrate intellectual history effectively with social, cultural and political history, then, it seems likely that what is really needed is nothing like 'cultural sociology' but rather a new, reformed intellectual history presiding over a two-way traffic, or dialectic of ideas and social reality, and focusing less on finished theories [. . .] than the 'reconstruction of polemical, frequently unresolved arguments', a new intellectual history in which the major theorists of the past still figure prominently, though the ranking between them may be greatly altered, but in which the chief emphasis is less on thinkers and theories than on 'thinking' and debates." [*ibid.*, 23]

Universalism and the challenge for Enlightenment reason

If there exists a class of qualities that deserve the name of cardinal intellectual virtues of Enlightenment the following ones, which Isaiah Berlin found the most representative, are bound to be included: intellectual power, honesty, lucidity, courage and disinterested love of truth. [Berlin (1956), 29] Accommodating these virtues under its moral firmament, the eighteenth century was pointing the way towards the future, in the hope of establishing a universal, or cosmopolitan republic of letters. The characteristic princi-

be judged on the basis of genetic analysis. A case in point is Cassirer's multi-volume study on the history of the problem of knowledge (*Das Erkenntnisproblem in der Philosophie und Wissenschaft der neueren Zeit.* (Band 1: 1906; Band 2: 1907; Band 3: *Die nachkantischen Systeme*, 1920; Band 4: *Von Hegels Tod bis zur Gegenwart (1832–1932)*, 1957) [Cassirer (1950)]) which still is one of the best accounts of the history of ideas in the modern era ple that formed its central ideological core was the notion of the *sovereignty of reason*. Indeed, the radical *philosophes* of the Early Enlightenment – Bayle, Fontenelle, Boulainvilliers, Meslier, Fréret, Boreau-Deslanders, Tyssot de Patot, Du Marsais, Rousset de Missy, La Beaumelle, Lévesque de Burigny, Mably, Morelly, Diderot, d'Alembert, Helvetius, the marquis d'Argens and the pre-1754 Rousseau – contended that reason teaches that human society should be based on personal liberty, equality, and freedom of thought and expression. In the atmosphere enlivened by these ideals alone, it is understandable that science and philosophy experienced their greatest victories so far in history. In the cultivation of the "enlightenment virtues", the great thinkers and teachers of the eighteenth century remain surely unparalleled. However, these virtues presupposed a *forma mentis* that was to pose the greatest challenge to the ideal of Enlightenment rationality in the following historical era. This mode of thought is closely related to the idea of universalism.

We have been gradually unfolding the elements out of which the idea of universalism eventually emerges as a full-fledged regulative principle of both scientific and ethical thought. Some of these elements were crucial, not only in bringing forth more clearly the underlying idea of universalism, but also in bringing about circumstances in which the Enlightenment idea of rationality was challenged from two sides: from within the Enlightenment ideas themselves as well as from outside through the ideas of counter-enlightenment. The characteristic feature of Enlightenment thought which organically bound it to the idea of universalism and exposed it to the two-sided challenge was the idea of criticism. Not only historians have characterized the Enlightenment in these terms, but contemporaries also envisaged their times as the 'age of criticism'. Immanuel Kant defined the essential characteristics of the age in the preface to the first edition of his *Kritik der Reinen Vernunft*:

> Our age is, to a pre-eminent degree, the age of criticism, and to criticism everything must submit. Religion through its sanctity, and the state through its majesty, may seek to exempt themselves from it. But they arouse just suspicion against themselves, and cannot claim the sincere respect which reason gives only to that which sustains the test of free and open examination. [A xii]

However critical the Enlightenment *philosophes* strove to be in their intellectual analyses of man, state, religion and science, they could not escape some severe philosophical problems that undermined the plausibility of their ideal of rationality. In addressing the relation of the general outlook of the logical empiricists to the ideals of Enlightenment André Carus presents two fundamental problems that the Enlightenment thought was not able to solve satisfactorily (thereby setting a challenge to modern philosophy and science which the members of the Vienna Circle found to define the content of their overall philosophical agenda). The main problems were, according to Carus, (1) the first obstacle to Enlightenment: no criterion for knowledge, (2) the second obstacle to Enlightenment: knowledge and the practical realm. [Carus (2007), 12–31] What do these problems actually consist in, and what is their relation with the tradition of Enlightenment in particular?

¶The first question boils down to this: What *is* the 'knowledge' that the Enlightenment regarded as so critical to individual autonomy and social improvement? Can it be defined in practical terms? What were the criteria that the *philosophes* set for knowledge, i.e., what *qualifies* as knowledge? Furthermore, given the compendium of everything that qualifies as knowledge, *how* does it fit together? This problem is very important, given that one of the most important methodological motives of Enlightenment is to build a *system* of knowledge. Thus, we are interested in the principles that provide a basis for cognitive systematization. In addition to these questions, one can finally ask: what constitutes *important* knowledge? It is this question that posed insuperable obstacles to the Enlightenment philosophers.

¶The second question concerns the tension created by the apparent blindness of *scientific* (and especially theoretical) knowledge – which occupied an exemplary place in the Enlightenment canon – to the *moral* and *affective* context of knowledge. An often presented complaint among the Romantics was that giving science such priority as the Enlightenment outlook supposed would degrade our subjective perceptions to the status of 'mere appearances', while impersonal scientific formulas specify the 'underlying reality'.

How did these challenges shape the philosophical atmosphere of the Enlightenment? What were the reasons for the failure of the Enlightenment philosophers in addressing these challenges? In Carus's view:

> The Lockean or Baconian empiricism of the *Encyclopèdistes* could not provide an adequate criterion of knowledge for the [following] reason [. . .]: it was unable to account for mathematics, or for the central importance of mathematics in scientific knowledge. Moreover, it put the burden of fitting all of knowledge together on a speculative taxonomy of human cognitive faculties or on the mechanical programme, which had to be accounted equally speculative as it could not be applied to any actual knowledge. The effort to overcome these problems by Enlightenment-minded thinkers over the next century took two main forms: the 'positivist' tradition following Comte and Mill, and the Kantian tradition. [*ibid., 12*]

Carus implies that the essential difficulty in the proposed ideal of knowledge in the Enlightenment tradition was the impossibility of accounting for the abstract and universal nature of mathematical knowledge with the conceptual tools and presuppositions of the empiricist tradition on which the main representatives of Enlightenment philosophy leaned on. Although the attempts at meeting the difficulty made by Comte and Mill on the one hand, and Kant on the other, were genuine improvements over the pursuits of Enlightenment philosophers, they nevertheless fell short of their attested objectives. Even Kant, who is suggested to have come closest to an adequate solution, failed because of a fundamental dualism that was built into his system. Indeed, Carus evaluates Kant's contribution in this respect as follows: "Whatever his intentions, Kant's legacy in the nineteenth century (especially in German idealism and historicism) had been to split knowledge into two sectors: the natural sciences, based on a priori mathematical principles and justified by perceptual evidence, and the human sciences (*Geisteswissenschaften*) based on entirely different principles, such as the freedom of human action and the existence of cultures or communities of shared action-guiding values. The enlightenment emphasis on the *unity* of knowledge – the cognitive as the *conscience* of the practical – was thus lost." [*ibid.*, 12–13] The whole purpose of Carus' discussion is to set the following idea in a historical context: it is through the combined efforts of the logical empiricists, and particularly Rudolf Carnap's, in the early twentieth century that these Enlightenment problems are finally solved *in tandem*. This solution is made possible by the conceptual tools of the 'new' mathematical logic and the completely novel conception of philosophy as a form of 'language engineering' (the roots of which Carus traces to the "engineering and revolutionary spirit of the Ecole Polytechnique in the early nineteenth century [. . .] and was elaborated in the works of Condorcet, Saint-Simon, and especially Comte." [*ibid.*, 14]). The gist of Carus' view is that motivated by Enlightenment ideals, the logical empiricists, and foremost Carnap, are finally able to meet the objections presented against the classical conception of knowledge and its value that underpinned the systems of Locke, Kant and Comte, for example. For the first time in the history of thought, the two problems of knowledge can be adequately addressed and solved!

Although I agree with many of Carus' views about the unique character of Carnap's solution, I hesitate to render it the philosophical and ideological status that Carus does. Indeed, it is precisely because Carus fails to see the *difference* and delicate relationship between the domains of formal-universal laws and axiological validity that his reconstruction of Carnap's solution is *not* a solution of the two-fold problem of Enlightenment thought. The idea of universalism that Carus alludes to with the locution "unity of knowledge" is only a limited variety of universalism, universalism solely in the sense of formal-universal laws, that seeks to conflate the domains of fact and value. This is, as we have already shown above, a philosophical mistake. What Carus' account precisely lacks is the explicit reference to the all-important influence of the tradition of intellectual virtues upon the gradual evolution of the problem of knowledge, and more generally, the regulating role of the ideal of universal validity in determining the relationship of different values to each other. It is not solely through formal considerations such as dwelling on the indisputably deductive nature of mathematical knowledge that the background of the problem is set up. The main point is, as I believe Lorraine Daston and Peter Galison have convincingly shown [Daston & Galison (2007), *passim*], that the intellectual virtues (such as *truth-to-nature*, *objectivity* and *trained judgement*) constitute important boundary conditions for addressing the problem of knowledge in the first place. The *ethico-epistemological* virtues and the conception of knowledge are (and have been throughout the history of science) in a dialectical relationship where the two-way traffic between them continuously changes the conceptual landscape in which the problem of knowledge is addressed. Central to the examination of the universalistic ideas about moral virtues and knowledge is the idea of epistemic values associated with the concept of *scientific self*. In the next section I will examine the notion of *scientific self* in its relationship with objectivity and the ideal of universalism.

2.3 Universalism and the Idea of *Scientific Self*

2.3.1 Epistemic virtues and pre-social virtues

The idea that reference to epistemic virtues is not solely a metaphorical *façon de parler* but an indication that they are literally *moral* virtues can be seen to derive already from Plato and Aristotle. Although their ethical views were very different, they both shared the idea that human dignity consists in the fact that a man must make himself what he ought to be. The drama of human life is reflected in the fact that a man can either fail or succeed in doing so. Virtues and vices, great and small, are the prerogatives of humans and the boundary conditions of that inalienable dignity that belongs to each person; man can either choose to act virtuously or let his actions be determined by vices. If we define virtue as a disposition (or trait) to promote human good and flourishing, then we have several types of virtues (behavioural dispositions) depending on which part of human life they relate to. Indeed, from the first-person/third-person point of view there definitely can be said to exist both pre-social and social virtues, because moral is not exclusively a group phenomenon.[14] The term pre-social is misleading in so far as all virtues, whether pre-social or social, are important for the whole community and can develop in it. I will nevertheless stick with this terminology, because it highlights the importance of the individual dimension of morality that is relevant for the notion of self in general, and *scientific self* in particular.

What is the actual constitution of the system of virtues and vices? How are they related to the nature of man? In order to make sense of the notion of epistemic virtue, we have to place it in the larger theoretical structure of virtues *simpliciter*. This was seen already by Plato. For example, regarding wisdom, one of Plato's cardinal virtues, as an epistemic virtue, already presupposes a total system of virtues, because Plato identified (at least) six different forms of wisdom: (1) *craft wisdom*, as in the wise physician, (2) *personal wisdom*, as codified in the maxim "Know thyself" (γνῶθι σεαυτόν), (3) *the wisdom of counsel*, (4) *social wisdom*, (5) *theoretical or esoteric wisdom*, and (6) *wisdom as the meta-virtue*, the virtue which rules the other virtues. Following the platonic idea of a system of virtues (an idea which Aristotle adopted in his *Nicomachean Ethics*), we provide a brief synopsis of the pre-social virtues that underpin the idea of epistemic virtue.

The system of pre-social virtues

From a general point of view, *life* is undoubtedly the basis of personality, and person's life taken as the totality of conditions for self-actualization is the most elementary good.[15] Indeed, if the person's life is destroyed, he can enjoy no other goods. But then, in the hierarchy of goods life constitutes only a *basic* good and there are goods that are qualitatively *higher* than the purely vital values. The significance of life as a value derives from the

[14] The distinction is due to Hösle [Hösle (2004), 281].

[15] In this section I follow the lucid presentation of Vittorio Hösle [Hösle (2004), 281–304].

ontogenetic fact that only through it can forms of life and *Lebensgefühle* develop that transcend the vital *per se* and allow the emergence of that which is truly human in human beings, i.e., that which transcends the state of being of humans as mere organisms. There is a peculiar asymmetry in the degree of the *sanctity* of life depending on whether the life of a person is examined from the point of view of others or the person himself; juridical systems around the world almost universally sanction *others'* attacks on a person's life as the greatest crime, but from the point of view of the person, risking *his* own life may be a personal duty that enables him to become fully worthy of it in the highest, moral, sense. The primacy of vital values is manifest in law, where all other goods are subordinate to life.[16] From a personal point of view, on the other hand, life can be considered fulfilled only if in the course of it the person is able to complete its process of self-actualization, in short, if the *I* is able to realize its self.[17] What, then, is the character of the higher good that alone conveys life its meaning? Although we have to forego here a fully elaborated metaphysics of man (comprising in essence an answer to the question "What is man?"), it is pertinent to provide some hints as to what the highest coordinating principle of morality could be. One answer that makes the distinction between basic goods and higher goods in the most candid manner is provided by Vittorio Hösle:

> The *spirit* is what first gives *life* – the basis of spirit – its true meaning. In ethics, the complex relationship between life and intellect – which is so complex because life is the *real* presupposition of intellect, and the intellect the *ideal* presupposition of life – results in many difficulties that would not arise if we had a one-dimensional hierarchy of values; however, this mutuality, which is grounded in the dialectic of the ideal and the real, contributes a lot to the intellectually fascinating beauty of the moral. Even if not every person is asked to risk his life, relinquishing for the sake of intellectual values some of one's own vital values, which are bound up with the self-enjoyment of the living being qua living being, is profoundly human. [Hösle (2004), 281]

It can be asserted that the precedence of intellectual over vital values is a central norm for any individual person, even if one does not subscribe to the idealistic articulation of the distinction given by Hösle. For clearly, there is a sense in which the idea of the priority of intellectual values can be rendered intelligible in the social context as well. For example, a scientific theory or theoretical insight concerning a specific domain can be shared and utilized by several people without any loss of information value, provided that everyone is sufficiently prepared and skilled enough to benefit from the theory (or insight). On the other hand, one cannot share a physical source in the same manner, for example a cake cannot be eaten whole by several people. Furthermore, even a book which contains a description of the theory in question cannot be used by several people at the same time.[18] Therefore, the distinction between *abundant* and *scarce* resources

[16] These other goods include: (i) the *inviolability of the body* and formal self-determination; (ii) *property and fortune* through which a person can carry out his more complex plans and (iii) *personal honour* as the medium in which a person must operate socially.

[17] I will touch upon the human identity problem and its relevance for the dimension of morality and moral action in the Chapter **3**.

[18] These examples are due to Hösle. [*ibid.*]

constitutes a demarcation between goods that can be shared freely, allocating basically similar utilities to everyone involved, i.e., ideally the same utility to everyone regardless of the fact how many people are sharing the good, and goods that cannot be so shared. The goods belonging to the latter category are the subject of the theoretical sphere of economics. To reiterate this distinction and its significance for the notion of the social, Hösle writes: "The noema of Beethoven's Ninth Symphony rises above the circle of economics, but the individual concert performances of it, which cannot be heard by everyone, do not. The intellectual person is thus, *ceteris paribus*, more social than one who pursues material pleasures; if two people are interested in the same intellectual insight, the latter constitutes a kind of bond; if, on the other hand, two people would like to have the same piece of land, conflicts are likely to arise. And yet the term 'pre-social' is meaningful: it is intended to suggest that, in contrast to social virtues, these virtues could exist even if only a single person survived." [*ibid.*, 282]

The first of the pre-social virtues in the system of virtues that we are here trying to explicate and that underlies the notion of epistemic virtues, is *temperance*.[19] We mean here the notion of temperance in its broadest sense that refers to the ability to forego immediate satisfaction of needs, whether these are of a general biological nature or specifically human. Certain reservations are, however, in order here. Although one could use the demarcation between biological and genuinely human needs to sort out varieties of temperance, referring in the first case to moderation and in the second case to temperance in a narrower sense, there are at least two necessary qualifications to be made with respect to temperance as a virtue. First, the distinction between speciously virtuous individuals and genuinely virtuous individuals can explain the skepticism with which some regard the virtue of temperance. Some people may well exemplify remarkable temperance which objectively amounts to little more than resisting pleasure for the sake of some illusory good. Those who have made this virtue their pride seldom realize the functional role of temperance within the system of virtues. Indeed, as Hösle has remarked, "[e]ven if moderation can be expected if everyone, it rises to *asceticism* in only a few – although admiration for these view is general. Even an ascetic can be intemperate in a narrower sense of the word: if ascetic behaviour becomes compulsive, if the ascetic is not able to limit his habits when the circumstances demand that he do so, a loss of autonomy occurs. [...] To torment oneself pointlessly is in any case contrary to duty. Asceticism is most meaningful when vital values are sacrificed to intellectual values." [*ibid.*, 284] Needless to say, (false) pride in the exercise of a virtue usually is a mark of a speciously virtuous individual, and this generalizes to the exercise of all virtues, not only temperance. Genuine or true pride is intrinsic, having its measure in the individual himself, and is thus not in need of others' homage. Second, some cause of conceptual confusion may well have resulted from the imprecise translation of Aristotle's σωφροσύνη and Aquinas' *temperantia*, which have been conventionally translated as temperance. This is quite unfortunate because serious philosophical misunderstandings supervene here on philological inaccuracies. The most important of these is the conception that temperance is (philosophically) synonymous to both moderation and self-control. This is, in broad outline, also Hösle's

[19] I will largely draw on Hösle's inclusive classification.

view.[20] It was both Aristotle's and Aquinas' conviction that moderation and self-control are necessary, but not sufficient, conditions for temperance (while it can definitely be said that they constitute the core intension of temperance[21]). Whereas moderation is usually associated with asceticism and (unnatural) suppression (along the lines of Hösle's definition), temperance should be associated with reason, harmony and human flourishing. Furthermore, Aristotle does not identify self-control, or continence, with temperance in its genuine sense. For Aristotle, continence and incontinence do not yet qualify as virtue and vice respectively. This is a stringent condition; for Aristotle only habitual forms of behaviour that have been deeply ingrained in the individual's character through systematic practise can be regarded as genuine expressions of virtue.[22] On this account, an incontinent person can *perform* a virtuous act – in most cases only through an experience of inner conflict, willing to control his appetites – but cannot *be* virtuous. Of course, such experiences of inner conflict and overcoming the personal weaknesses through genuine acts of will are necessary for obtaining moral autonomy; they constitute acts of self-abnegation through which *I*-ness is first constituted. Only through such practise can persons become "unified in motivation and deliberation" which is a true mark of a virtuous person. Temperance, which usually is concerned with the regulation of behaviour related to human metabolism and reproductivity, has been enculturated in all societies throughout history. As Hösle remarks,

> Notwithstanding all the differences between individual cultures, there is no society that did not in some way transform metabolism and especially reproductivity. Among humans, eating and drinking are not mere biological processes,[23] and hence in German even their names differ from those of the corresponding processes among animals. Hunger is not the only thing that may determine when we will eat; for instance, sacrifice before eating is an expression of forbearance whose manifest function is a propitiation of the divinity, but it is in reality an end in itself, to be precise, a confirmation of the partial autonomy of humans with regard to their drives – a confirmation that is also served, for instance, by periodic fasts. Much the same holds for the sex drive. [Hösle (2004), 283]

Aristotle and Aquinas, the two architects of the virtue of temperance, do not reject food, alcohol and sex as intrinsically bad, something that would be beneath humans as rational and spiritual beings. Rather, they invite us to appreciate the goodness of these pleasures rationally and moderately. The significance of temperance consists in the fact that it is not

[20] Nothing very serious hinges on this, however, as Hösle's classification of virtues includes higher virtues the characteristics of which coincide with the Aristotelean conception of temperance. The important point is that no single virtue can be fully developed and exercised in isolation from the others.

[21] The core intension of a concept can be defined as $I_{\text{core}}(C) = \{P_1, P_2 \ldots P_n\}$, where the $P_1, P_2 \ldots P_n$ are the relations and properties subsumed under the concept. They are necessary and sufficient for the unequivocal distinction between C and any other concept, but do not constitute a complete characterization of C.

[22] The intentionalism of Hösle's ethical theory is in consonance with this aspect of Aristotle's ethics.

[23] Cf. Sidney W. Mintz & Christine M. Du Bois (2002): "The Anthropology of Food and Eating", Annu. Rev. Anthropol. 2002. 31: 99–119: "Like all culturally defined material substances used in the creation and maintenance of social relationships, food serves both to solidify group membership and to set groups apart."

oppressive but liberating in that it is necessary not only for the health of body and mind, but also for εὐδαιμονία, that is, (Aristotelean) happiness or human flourishing. Aquinas is absolutely right in his statement of the initially paradoxical contention that it is only through temperance that pleasures can be truly appreciated. Indeed, controlling drives can be said to have intrinsic value, because it increases the power of personhood. [*ibid.*]

In connection with epistemic virtues, temperance in a broader sense may be defined to include the systematic training of thought. One might here consider the Habermasian domains of formal rationality – (1) cognitive-instrumental reason; (2) moral-practical reason; and (3) aesthetic-expressive reason – in order to pinpoint the nature and scope of the thinking capacity that is being trained, but only with the proviso that it is not solely an instrumental reason that we are here concerned with in the sense that temperance is needed in order to perform some other tasks that are morally obligatory. The criterion of objectivity is a significant regulative principle also with regard to temperance: the effort to be objective in one's knowledge acquisition tasks, broadly conceived in the sense of attempting to grasp the structures underlying reality, "what truly is", definitely has intrinsic value. There is also a distinction with respect to the ability to know the universal, for example the (formal-universal) laws relating to a certain phenomenon, and the ability of concrete subsumption; it is intrinsically more valuable to know the general mechanism of a disease, for example the demonstration of the nature of the genetic defects in cystic fibrosis, than to be able to apply this knowledge to a single case;[24] the ability of concrete subsumption is temperance in a narrower sense. Furthermore, there is a sense in which one can evaluate temperance also on the basis of the value of the thing known; if one rises above the examination of concrete sphere of empirical reality to the sphere of (formal-universal) laws, as in the contemplation of the mathematical structure of a theory or the conceptual content of a theoretical insight, one is involved in activity that has high intrinsic value in itself especially if in so doing one grasps the value of his (inner) activity and realizes that the activity of reflection has become for him an end in itself. A well-known example of this distinction is the traditional division between pure and applied mathematics; for a theoretical spirit (as well as for a universalist) it is more noble *ceteris paribus* to seek to acquire mathematical knowledge for its own sake, than to seek it only in order to increase the sphere of utilities. The famous British mathematician G.H. Hardy used to refer to practical applications of mathematics as "glorified plumbing", and did not hasten to add that "[j]udged by all practical standards, the value of my mathematical life is nil . . . ." However indifferent this sounds from the point of view of human action and poietic virtues, it nonetheless brings forth the contrast between a contemplative and prudent attitude towards theoretical knowledge; Hardy was still deeply ingrained with the idea that the knowledge of the universal has intrinsic value in itself.

Philosophically it is interesting to note that the characteristics of *wisdom* – here defined as a pre-social virtue – have been taken by classical philosophers as the basis for determining the scope of philosophy. Although the philosophers of the Antiquity from Plato on took σοφία as the qualifying characteristic of the σοφός towards which the φιλόσοφος aspired, the reading of *wisdom* as the ability to interpret the individual on the basis of

[24] This is not to say that a realization of concrete goals in reality, e.g. treating patients efficiently, does not have intrinsic value; the contrast here is epistemic.

the whole emerged only gradually in Western thought. R.G. Collingwood set the task of philosophy precisely in terms of holism; as we saw in Chapter **1**, according to him philosophy is first and foremost an attempt to grasp a subject in its totality and in terms of its relation to the whole. For an individual, the possession of wisdom means that one is able to envision one's life and the rich textures of the experienced *Lebenswelt* as a whole, extending the attitude of wholeness also to one's own misfortunes which thereby can be understood and accepted in their necessity and taken as a possibility for personal growth. As Hösle beautifully puts it, "there is no purer and more unconquerable source of happiness, that is, of reconciliation with the world, than wisdom." [*ibid.*, 285] From a historical point of view it may be noted that it was mainly in the context of religious life that the virtues of *ascesis* and *contemplation* were rehearsed and in which they achieved a qualitatively superior state of refinement with respect to the common ethical life of the early cultures. The fundamental division between people who possessed these virtues to a prominent degree and those who did not was the basis for a religious subsystem that conferred positions of power and leadership to the individuals in whom these virtues were associated through religious training. Although the prudential aspect of the exercise of these two virtues still plays a role in contemporary religious communities, a major evolutionary step was taken when prudence was differentiated into the subsystem of science which then completely emancipated itself from religion.

We have already alluded to the domain of human action and its significance for the moral. Vittorio Hösle calls *poetic virtues* those that have to do with the realization of goals in reality. There are three kinds of poetic virtues, all of which are relevant for the framework of human action; they are historically determined by the tradition of craftmanship under whose purview they once were united, and only later differentiated to the characteristic virtues of modern science, technology and art. The tradition of craftmanship is important also in the additional sense of having provided the motivation for the philosophical analyses of 'poetic rationality' in the early modern philosophical literature.[25] What, then, are the three poetic virtues, and what does their significance consist in? The importance of action for the domain of the moral is clearly outlined by Hösle along with the characteristics of the three virtues:

> Not only knowledge, but also action ennobles man; imposing one's own goals on nature is a sign of the elevation of the intellect over nature and should be considered positive if the values of nature [...] are not violated. The latter is in no way necessarily the case; traditional agriculture even increased the diversity of species. Although intellectual values rank higher than vital values, the intellectual man must also live in actual reality: satisfying moderate needs through one's own industrious labour, bringing production and consumption

[25]The history of the inception of the works of Descartes is a case in point. D. was in a close contact with many craftsmen and scholars of his time the most prominent of which were Isaac Beeckman (1588–1637), a Dutch philosopher and scientist who was considered to be one of the most educated men in Europe, and Johann Faulhaber (1580–1635), a German mathematician who was trained as a weaver and later took the role of a surveyor of the city of Ulm. He collaborated with Johannes Kepler and Ludolph van Ceulen. Besides his work on the fortifications of cities (notably Basel and Frankfurt), Faulhaber built water wheels in his home town and geometrical instruments for the military.

> into balance, making the most out of the least, being sparing with resources and at the same time operating without meanness in this subordinate sphere are all elements of *economic virtue*. The essence of *technological virtue* is disciplining oneself in such a way that one deals with the demands of the object and succeeds in becoming adequate to it, even creating new structures that did not previously exist and increasing the stock of reality by adding artifacts to it. If at the same time a sense for beauty is not lacking and the artifacts are given a form that is not merely useful but also refers through the sensible to ideal relationships, one speaks of *aesthetic virtue*. [*ibid.*]

The importance of aesthetic virtue for the person's life in general, but also for the scientist's life and work in particular, is an especially interesting issue which we can barely touch upon here. If one considers the historically evolved conceptions of scientific self that build on different epistemic virtues as the ideal constitution of a scientist's character, it is clear that the dimension of the aesthetic also plays a prominent role in the formation of a scientist. Important as this subject is, we cannot unfortunately delve into the matter in this dissertation; the subject would surely merit its own treatment in a separate treatise.[26] To relate the aesthetic dimension to the dimension of the epistemic, we can say that through its realization of inner values art is in its highest forms akin to contemplation and approaches the sphere of ideal validities, because like contemplation, it rises to the 'first world' but yet makes it accessible to the senses through material representations that everyone can appreciate.

There is one more pre-social virtue, one of the classical cardinal virtues, without which the classification would be incomplete. Whereas the temperate person foregoes, as we saw above, the satisfaction of a few vital needs to increase the power of personhood, he does not risk life itself in the pursuit of this control over himself. *Courage* is precisely the virtue that constitutes the readiness to put life at stake, the "actual advancing toward death." [Hösle, 285] The relation between these two virtues is interesting, as already Plato noted. As Hösle writes, "[o]n one hand, temperance and courage both involve an overcoming of immediate vital instincts; on the other, the instincts that are overcome are very different: In the case of temperance it is a matter of pleasure, and in that of courage it is a matter of fear – in its early forms, the fear of disadvantages in general and, in

[26]Another topic related to the aesthetic dimension of science is the notion of intuition in the context of scientific discovery. Although philosophers of science have been conditioned to be critical of the notion of intuition, it is a sociological fact that many scientists – most prominently mathematicians – constantly refer to their intuitions as an important vehicle for discovery and scientific communication. For example, Alexandre Grothendieck (1928–), the central figure behind the creation of the modern theory of algebraic geometry, relates how he was deeply impressed by Jean-Pierre Serre's (1926–) ability to communicate something akin to intuition: "The essential thing was that Serre each time strongly sensed the rich meaning behind a statement that, on the page, would doubtless have left me neither hot nor cold-and that he could 'transmit' this perception of a rich, tangible, and mysterious substance – this perception that is at the same time the desire to understand this substance, to penetrate it." [*Récoltes et Semailles*, 556] Of course, it is a completely different matter to explicate the precise cognitive processes and mechanisms behind such intuitions, but the task of explication is not facilitated in the least by the kind of negligence that is shown towards the myriad phenomenological reports of intuition among scientists. A *phenomenology* of creativity remains a desideratum within the psychology of creativity.

its most advanced form, fear of death [...]" [*ibid.*] The polar opposition of the virtues of temperance and courage can be put succinctly: temperance inhibits, courage sets in motion. The scope of courage is not, of course, restricted solely to the defiance that one exhibits upon facing death or natural forces; it can equally be exercised in contexts where one's acts of courage are directed against enemies of the group or one's own group. The scale of difficulty and, one would be inclined to say, nobility with respect to the varieties of courage runs from the battle with the natural forces through the conflict with foreign conspecifics to the opposing of those to whom one is bound by collective identity.[27] Although the latter does not often involve any risk to body and life, but only social isolation, it is most difficult precisely because it demands what is aptly called "*the courage of one's convictions*" within a group which has played a prominent role in the formation of one's self-conception. Therefore the self-renunciation required in defending a new theory in a hostile scientific community, for example, is a mark of greater courage than the battle to survive in the severe physical circumstances of a mountain range. However, this form of courage is not inclusive and, as Hösle correctly annunciates, "neither is it possessed by every person who does not shy away from death at the hands of an enemy, nor need a person who is endowed with it be able to look death in the eye without blinking." [*ibid.*, 286] The "courage of one's convictions" figures prominently in many scientific debates and it comprises a genuine mark of a developed self-conception of a scientist. We will come across examples where this kind of courage has a peculiarly important role, especially in pointing the direction for future research in disciplines where the established forms of thought are unable to solve some outstanding open problems. In the following section we will see how epistemic virtues that constitute the background framework for the cognitive and moral conduct of a scientist, are related to the overall system of pre-social virtues we have sketched above.

2.3.2 The notion of *scientific self* within the history of science

As we mentioned in the preceding section, the idea of epistemic virtues was already central to the classical philosophers from Plato and Aristotle on. In the Enlightenment the idea of epistemic virtue was developed further in the context of natural philosophy when the *philosophes*, who took up the science of the preceding age and helped to establish it as the dominant force in the Western societies, started to ponder over what it was that made science so successful in describing the world and producing technical applications on the basis of the knowledge of the formal laws of nature. The insight that it was not so much the science itself but the new method, as rehearsed by Bacon, Descartes and Newton, which seemed so potent and successful, prepared the way for the concrete evaluations of *practices that were conducive to truth*, i.e., *epistemic virtues*.

[27]The necessary conditions for collective identity are that human beings sharing a collective identity all have (1) common contents of consciousness; they must (2) know that they have common contents of consciousness, and they must even (3) know that all know that they have common contents of consciousness. These requirements resemble the conditions for common knowledge provided in the literature on microeconomics. The concept was first introduced in the philosophical literature by David Kellogg Lewis in his study *Convention* (1969).

What, then, are the so-called epistemic virtues? Daston and Galison define them as "norms that are internalized and enforced by appeal to ethical values, as well as to pragmatic efficacy in securing knowledge". [Daston & Galison (2007), 40] Here we have an explicit statement by respected modern historians and philosophers of science that epistemic virtues have their grounding in a system of ethical values, and this is certainly right from a historical point of view. However, the relationship of these values to the requirement of pragmatic efficacy is precisely the question of interest: which particular values underpin the epistemic practices of scientists and how do they facilitate pragmatic efficacy in securing knowledge? The strategy of Daston and Galison is to show that the epistemic virtues have a history – just as mores and morality have a history – and that different conceptions of epistemic virtue derive their legitimacy from the different requirements of pragmatic efficacy which are connected with the conception of knowledge that determines the possible methodology of science. In particular, the notion of objectivity the history of which Daston and Galison survey in considerable detail, is seen to be a fundamental coordinative concept around which the different ideas about epistemic virtues and the notion of scientific self are wound. One of the first ideas about objectivity in the early nineteenth century was that in order to represent a scientific object faithfully one had to eliminate all subjective interferences in the process of inquiry (we are speaking here about the experimental context of course): "to be objective is to aspire to knowledge that bears no trace of the knower – knowledge unmarked by prejudice or skill, fantasy or judgment, wishing or striving. Objectivity is blind sight, seeing without inference, interpretation, or intelligence" [*ibid.*, 17]

Up to the early nineteenth century, the dominant view about 'objectivity' had been something completely different. The epistemic virtues characteristic of the age of Enlightenment and eighteenth century were exemplified most clearly in the works of naturalists; in their survey of the botanical works of that era, Daston and Galison show that the objects which an eighteenth-century botanic atlas sought to represent was not an array of specimens, a comprehensive catalogue of individual plants – unique in their singularity – but, rather the ideal form, the *type* of each species. The preparation of such atlases was concomitant to the attempts to "tame Nature's variability". These projects found expression, *inter alia*, in the intellectual and practical conduct of the makers of such atlases who, whether they were interested in things as diverse as plants, bones or crystals, shared a common aspiration to see past the surfaces of their objects of investigation and grasp their underlying forms. In order to grasp such a primary form, the naturalist was figuratively involved in a process of 'platonic purification': they created strictly regulated forms and techniques of seeing to produce images that best would represent "what truly is". Indeed, through tenacious and tireless observation, the naturalists strove to discard everything they judged inessential and accidental; in order to extract the universal from the particular one had to subjugate oneself to strict rules of conduct that required regular practise and quieting of the will. It is evident that this kind of protocol for scientific observation was far from passive; the naturalist had to exercise his intelligence and willpower actively in order to abstract the perfect form from the chaos of multiplicity. Quite appropriately, Daston and Galison call this particular epistemic virtue *truth-to-nature*.

One particular influence that cannot be passed over in silence in an examination of the formative currents in the development of the notion of *truth-to-nature* is the formidable impact of the personality of Goethe. His influence manifested itself most clearly in association with the contrast between the classical idea of culture conceived as a self-contained whole and the clear breaks and discontinuities in the understanding of culture with respect to nature. In a sense, Goethe's *Weltanschauung* was a reflection of his overarching attempt to reconcile these conflicting aspects of German culture. This conflict was most forcefully experienced in the transition from the nineteenth to the twentieth century. In Germany, the concept of nature has often been included in the concept of culture. Kant had been talking about a natural plan, to which the development of the human species was subject. The aim of development was a "state with a cosmopolitan purpose". For Herder, it is God's imprint in nature that can be deciphered from the development of culture, as if from a book. Goethe saw nature as culture's primordial image and model. For him the development of culture should be inspired by a deep and pure experience of nature, the essentials of which he once related in connection with his account of his discussion with Schiller. Goethe tells of a conversation that once unfolded between Schiller and himself after both had attended a meeting of the society of natural research in Jena. Schiller showed himself little satisfied with what had been presented in the meeting. A fragmented way of looking at nature had met him there and he remarked that such a way could not appeal at all to laymen. Goethe replied that it would perhaps remain strange even to the initiated themselves and that there could still be another way of presenting nature, not as something separated and isolated but rather as working and alive, as striving from the whole to the parts. He sketched "with many a characteristic pen-stroke, a symbolic plant" before Schiller's eyes. It was meant to show the successive becoming of the individual plant parts, their emerging from each other, and their relatedness to each other. About this symbolic plant shape Goethe, on April 17, 1787 in Palermo, wrote down the words, "There must after all be such a one! How would I otherwise know that this or that formation is a plant, if they were not all formed according to the same model." Schiller's reaction to Goethe's account of the basic ideas of his plant morphology was skeptical; he informed Goethe that the leaf he had symbolically sketched on paper was "no experience" [*keine Erfrahrung*] but "an idea" [*eine Idee*]. To this Goethe replied that "Then it is clear that I see my ideas with my eyes". [LA, **I**, 9, 81][28]This is one of the most beautiful illustrations of the unique character of Goethe's approach to natural phenomena; it is a paradigmatic exemplar of *truth-to-nature* coördinating scientific investigation.[29]

[28]The reference here is to the "Morphological Notebooks" of Goethe contained in the Leopoldina (LA) series. [Johann Wolfgang Goethe. *Die Schriften zur Naturwissenschaft*. Vollständige mit Erläuterungen versehene Ausgabe herausgegeben im Auftrage der Deutschen Akademie der Naturforscher Leopoldina. 2 pts, 11 vols. Ed. Rupprecht Matthaei, Wihelm Troll and K. Lothar Wolf. Weimar: Böhlau (1947–) Pt. **I**, Texte; Pt. **II**, Ergänzungen und Erläuterungen.]

[29]The novelty of Goethe's approach was perhaps most clearly exemplified in the realm of optics; the controversy between Newton's and Goethe's theory of colors is well known. Goethe's scientific project, hugely influential in the nineteenth century, epitomizes the influence of romanticism in science. To comprehend nature one has to abandon the abstract and fragmentary concepts of Newtonian science; instead a keen and pertinacious sense observation made possible by self-discipline and the habituated practice of quieting the will constitute a basis for a science that is simultaneously veridical and edifying. To make it explicit, for Goethe a world of ideas which does not permeate the things of nature, which does not bring forth their appearing and disappearing, their becoming and growing, is a powerless web of thoughts. The logical spin-

To return back to the issue of the general reception of the ideal of *truth-to-nature*, it was characteristic of the *forma mentis* of the early naturalists that the ideal engaged them in aesthetic and ontological judgements that were bound to general philosophical ideas then current in the intellectual circles. Through time this inevitably led to a situation where these judgements were questioned on the basis of a new notion of objectivity. However, later when the ideal of *objectivity* emerged as a new intellectual virtue (in the early nineteenth century as was mentioned above) the earlier ideal of *truth-to-nature* was not entirely suppressed. On the contrary, the problem of variability that had formed the point of departure for tackling the question about the possibility of scientific knowledge, remained an important reference point for later philosophies of nature as well. From a general historical vantage point it can be seen that the problem of variability/invariance persists throughout the development of science in the modern period, up until the twentieth century. Characteristic of this development is that the response to the problem of variability is dependent upon the intellectual (and ethical) conduct of scientists. This is succinctly pointed out by Daston and Galison: "[D]ifferent epistemic ways of life made for different diagnoses of the sources of variability. Eighteenth-century savants tended to locate variability in the objects themselves – in the accidental, the singular, the monstrous. By the mid-nineteenth century, the chief source of variability had shifted inward, to the multiple subjective viewpoints that shattered a single object into a kaleidoscope of images. The earlier naturalists had attempted actively to select and to shape both their objects and their illustrators, whereas later naturalists aspired to hands-off passivity. The meaning of the images changed accordingly." [Daston & Galison (2007), 113]

What were the next important steps in the development of the ideal of objectivity? For the German idealist philosophers, most prominently Fichte, Schelling and Hegel, objectivity was ontologically and epistemologically a product of thinking subjectivity (*ob-jectum*), a refined interpretation of the Parmenidean conception that "what is . . . is identical with the thought that recognizes it". Through the misinterpretations of Fichte's natural philosophy there arose a peculiar conception about the Fichtean "*I*" as antithetical to science. To counterbalance this pervasiveness and intrusiveness of the "*I*", the regulations of scientific conduct that guided the nineteenth-century scientist included commendations to be as passive and 'absent' as possible: this was taken to be a necessary condition for preparing circumstances in which the object could emerge in its purity. Obviously, an immediate consequence of this was that it would be best to let oneself be replaced by a machine. Machines capable of recording observational data (such as the photographic camera and phonograph, and later physical instruments such as the cloud chambers modeled after Aitken's dust chamber of 1888) were explicitly designed to record phenomena without human intervention, if not actually capable of accomplishing this. The link between objectivity and photography was, however, neither a necessary nor a sufficient condition for the emerging idea of mechanical objectivity; as Daston and Galison remark: "nonin-

ning out of lines of thought, without descending into the real life and creative activity of nature seem to him unfruitful. For he feels himself intimately intertwined with nature. This is an explicit rendering of the essentially Romantic spirit of Goethe's naturalism. His influence at the turn of the century was immense. In a sense he epitomized the inner yearning of those who, disappointed with the cold and rational outlook of life brought about and nourished by the industrial society, rehabilitated the romantic ideals that had held sway in the early nineteenth century.

tervention – not verisimilitude – lay at the heart of mechanical objectivity". [*ibid.*, 187]

The apogee of Daston's and Galison's narrative is reached when we come to the twentieth century; this is the epoch of *trained judgment*, the final member within the triad of cardinal epistemic virtues. Interestingly enough, the extreme forms of self-renunciation that were the hallmark of mechanical objectivity, led to a reaction in the beginning of twentieth century on the part of scientists who now insisted on the importance of intuition, judgment, and the capacity to interpret. These mental capacities were cultivated and the training of them was encouraged equally in the processes of knowledge acquisition leading to scientific discoveries and in the criteria that regulate the production of scientific images. Apart from the modes of inquiry that were characteristic of the earlier epistemic virtues – looking for the ideal forms hidden under the "Veil of Isis", i.e. the formal-universal regularities behind appearances (*truth-to-nature*), or quieting their own will and sacrificing their capacity for judgement (*objectivity*) – scientists now aspired to something that might be called "*physiognomic sight*": "a capacity of both maker and user of atlas images to synthesize, highlight, and grasp relationships in ways that are not reducible to mechanical procedure, as in the recognition of family resemblance" [*ibid.*, 314]. The 'natural' and 'simple' image, such as a photograph of a singular and unique scientific object (a crystal, a snowflake, a plant, an animal, a human organ), that played a prominent methodological and technical role under the purview of objectivity, turned out to be as enigmatic and impenetrable as the nature it promised to represent [*ibid.*, 357]; scientists soon acquiesced to the necessity of "reasoned images" that were drawn by an expert with a trained eye, and that were designed to train and instruct other eyes.

The above sketch of the development of the three cardinal epistemic virtues of scientific practise provides some important insights into the dynamics of the self-formation of the scientist. As was remarked in the beginning of this section, the epistemic virtues were internalized and enforced by appeal to ethical values and pragmatic efficacy. The issue of pragmatic efficacy has already been touched upon. We now turn to the question: what are the ethical values that underpin this conception of the development of epistemic virtues? The question is, then, what principles shaped the "epistemic ways of life" that in their turn determined the stance towards the problem of variability? It seems evident that the principles that are at play here are genuinely *ethico-epistemological* in the sense that the conduct of a scientist in the interrogative context in which he attempts to acquire knowledge is dictated by the goals (securing knowledge, grasping universal truths, formation of justified (true) beliefs, etc.) he sets for himself. The ethical dimension of the situation involves two distinct but interrelated aspects: (i) the prescribed goals of the investigator that derive from explicitly axiological considerations ("What do I find valuable or worthy of pursuing?") and (ii) the consideration of the appropriate means of attaining these goals ("What do I have to do in order to realize the goals I have set myself?"), including the personal requirements set for the investigator ("What particular characteristics do I have to possess in order to realize these goals?"). These aspects reflect the extent to which the character of the investigator is entwined with the interrogative process. It is precisely the inextricable link between the ethical and epistemological dimensions of enquiry that made the eighteenth-century *philosophes* and naturalists as well as nineteenth-century scientists aware of the need to cultivate what could be called a *scientific self*. The severe demands of scientific practice challenged scientists to develop in themselves qualities that are generally associated with virtuous men in the traditional sense. They had to learn to educate their wills to develop in themselves the necessary skills of "training the senses in scientific observation, keeping lab notebooks, drawing specimens, habitually monitoring one's own beliefs and hypotheses, quieting the will and channeling the attention". [*ibid.*] These different skills exemplify what Michel Foucault has aptly described as "techniques of the self" [Foucault (1989), 134] They are practices of the mind and body (most often the two in tandem) that mold and maintain a certain kind of self. The 'self' which manifests itself in these practices during the period under consideration is not uniform. Unlike Foucault, we can here see a veritable menagerie of different 'selves', both artistic and scientific, that are developed side by side, often developing in diametrically opposed ways. It is this pluralism of the techniques of the self – even within the purview of natural science – that calls for a systematic way to conduct scientific dialogue between various points of view.

The focal points of the discussion concerning "scientific self" are the concepts of *subjectivity* and *objectivity*. They constitute the background opposition against which the various conceptions of self, and especially of scientific self, can be evaluated. For example, regarding the different views on what comprises the essence of man's "being-in-the-world", the emphasis on subjectivity or objectivity dictates the way in which artistic and scientific selves are conceived and trained in the nineteenth-century. The interplay between subjective art and objective science, diametrically opposed as they were in the mid-nineteenth century, gave birth to a new conception of the role of subject in the process of observation: art and science ultimately converged in the dissolution of the self into its object. Nietzsche, for example, who was not a keen friend of scientific objectivity, nevertheless conceived a particular form of objectivity that was common to the best art and science: "There is required above all great artistic facility, creative vision, loving absorption in the empirical data, the capacity to imagine the further development of a given type — in any event objectivity is required, but as a positive quality. So often objectivity is only a phrase. Instead of the outwardly tranquil but inwardly flashing eye of the artist there is the affectation of tranquility; just as the lack of feeling and moral strength is accustomed to disguise itself as incisive coldness and detachment." [Nietzsche (1874) [Breazeale (1983), 93]] Nietzsche's proposal to put back together the two halves of the self – subjective and objective, active and passive, will and world – was counterproductive for a long time. It neither helped to reorient the practical conduct of the scientist in the laboratory nor to change the attitude of the artist depicting nature. The depth and extent of the problem was, however, now felt more fully. As Daston and Galison make clear:

> However illusory Nietzsche's "positive" objectivity may have been for both artists and scientists, it was proposed as a solution to a deep problem. Objectivity and the scientific self that practiced it were intrinsically unstable. Objectivity demanded that the self split into active experimenter and passive observer and that types of scientific objects be defined by atlas images of individual specimens too particularized to be typical. Nietzsche smelled the acrid odor of burnt sacrifice when the ascetic turned will against will: the objective

> man of science stood accused of inauthenticity, of self divided against itself. These were ethical reproaches. There were also epistemological objections to objectivity: How could an individual stand for a class without idealization or even selection? How could a universally valid working object be extracted from a particular depicted with all its flaws and accidents? [Daston & Galison (2007), 250]

It is clear that Nietzsche's criticism of the notion of scientific objectivity and the adjoining notion of scientific self helped to raise to general consciousness philosophical issues that had been passed over in silence in spite of their importance. Nevertheless, Nietzsche's acute observations on the particulars of the history of morality (not history of the moral, as the title of his famous book *Zur Genealogie der Moral* speciously suggests) could not, precisely because of his dismissal of the domain of ideal validity, make him appreciate the distinction between genesis and validity which is crucial for the constitution of the concept of the scientific self. We will now turn to study one prominent attempt to delineate the universal features of scientific self, i.e., that of Johann Gottlieb Fichte. In his work we can observe the seeds of the notion that the formation of a scientist is parallel with the moral formation of a self.

Fichte's *Vorlesungen über die Bestimmung des Gelehrten*

Fichte, who arrived in Jena in May 1794 after a decade of employment as an itinerant tutor, a decade marked by personal uncertainties, professional frustrations and constant financial insecurity, was determined to make a contribution to philosophy on a national scale with the intention of having a direct effect on the entire university community. To this end he enthusiastically seized the opportunity to realize his ambitions with the possibility of providing a course of "public lectures" that he was expected to deliver during his first semester. In spite of other engagements, most notably the elaboration of his systematic thought exposited in the *Wissenschaftslehre*, he was able to devote some time to preparing the lectures to be held in Reinhold's former lecture hall. By midspring he had his topic ready: "Morality for Scholars".[30] In the summary of the lecture series with which he began the first public lecture of the winter semester, a second round of the series (the first having initially commenced on May 23), Fichte spoke about the broad aims of his lectures:

> As most of you already know, the subject of my lectures is the scholar's vocation. The vocation of man as such is infinite self-improvement. This is also the final aim of all the social bonds between men. The *scholar's* vocation is to

[30] In the university catalog for the summer semester of 1794 the lectures are referred to with the different title *de officiis eruditorum* ("Concerning the Scholar's Duties") which was the title under which they were announced in the April 12, 1794 number of the *Allgemeine Literatur-Zeitung* ("Die Lehre von der Pflichten der Gelehrten").

> supervise this progress of culture in human society – to promote and to give it direction. [AA II: 3, 357][31]

The impact of the lectures on the young scholarly audience was immense and Fichte became a celebrity overnight. In a letter to his wife he relates how "people were standing on tables, benches and each others' heads". What was it that was so special in these lectures, and why do they still merit closer inspection two hundred years after they were first delivered? In the first place, the success of the lectures was in no small part due to the fact that Fichte was an eloquent and absorbing lecturer; as Breazeale remarks, "in both content and style, Fichte's public lectures were perfectly calculated to impress and overwhelm his audience. With their exquisite balance between confidential flattery of the audience, philosophical analysis, social criticism and moral exortation, these lectures bear striking witness to Fichte's public ambitions and talents." [Breazeale (1988), 142] In the second place, the lectures addressed a topical philosophical question that was intertwined with the most pressing social questions of the time. Similar considerations render Fichte's lectures relevant for our time and culture as well. Let us have a close look at these issues.

Although Fichte's lectures contain nothing that is conceptually obscure or unintelligible from the modern point of view, there are some issues concerning the title of the lectures that merit further scrutiny. We can concentrate on the two German terms *Bestimmung* and *Gelehrter* which present some difficulties for the translator. Let us consider the latter first. What is a *Gelehrter*? It is usually translated as "scholar" but a more proper translation would be "an educated person", for the intension of scholar is usually taken to be something of the sort "narrow professional occupation with texts and editions". Moreover, as Fichte stresses in his fourth lecture, a *Gelehrter* is not only an educated person; he is also one who "dedicates his life" to the acquisition of knowledge. In Breazele's words, "the true 'scholar' is not merely a researcher and teacher; he values knowledge precisely for its vital contribution to the advancement of mankind. Thus it is the special responsibility of the scholar to supervise and to regulate human progress towards perfection, and in order to do this he must at least strive to be 'the ethically best man of his time'." [*ibid.*, 141] How about *Bestimmung*, then? It is usually translated quite properly as "vocation", but again the English counterpart does have a significantly narrower meaning than the German one. If we look at the corresponding verb *bestimmen*, whose meanings are captured by "to specify" or "to determine", we see that *Bestimmung* then suggests that in the schema "*die Bestimmung* des *x*" we are talking about the "determination of *x*" or "the specific nature of *x*", and in a more general sense about its "characteristic" or "determining feature". [*ibid.*, 141n] Fichte is therefore interested in the question what it *means* to be an academic (in a broad sense).

Fichte's lectures are important precisely because they are not solely concerned with some special question about the vocation of academics but rather embed the discussion about

[31]The reference is to *J.G. Fichte: Gesamtausgabe des Bayerischen Akademie der Wissenschaften*, ed. Reinhard Lauth, Hans Jacob and Hand Gliwitsky (Stuttgart-Bad Cannstatt: Friedrich Frommann, 1964–) [Henceforth AA].

the duties of a scholar to a broader framework of philosophical anthropology; Fichte asks what is the nature or "vocation" of man as such. The discussion is inextricably linked with social and political philosophy. In the second lecture Fichte proposes an ingenious account of man's social nature and presents his idea about "a fundamental drive" toward society and social improvement. Only the fourth lecture is explicitly concerned with the question of "the scholar's vocation". I will mainly concentrate here on the theses developed in this lecture, because the ethical and social context of morality *per se* have already been extensively discussed above and Fichte's account of these matters does not significantly add to the ideas we have exposited earlier.

Fichte grounds his discussion of the scholar's vocation on the account of social classes he has developed in the earlier lectures. The equality of the different classes of the society is the *fundamentum* of his conception of the role of scholars and he stresses most adamantly that

> [...] every class is necessary and deserves our respect, that an individual is not determined by the class to which he belongs, but rather by the way he fulfills his role as a member of that class. For every person deserves to be honored only insofar as he *approximates to fulfilling his role completely*. For this reason, the scholar has reason to be the humblest person of all: since the goal which is set for him must always remain very distant, and since he has to achieve a very lofty ideal – one from which he normally remains very distant. [*ibid.*, 170 [AA VI, 324]][32]

This passage makes evident how deeply Ficte was influenced by the flourishing Enlightenment ideals (remember that the great French revolution had taken place just a few years earlier). The social dimension of his thought was determined by the idea that the division of labour is a natural requirement for a just society; "within society every individual quite legitimately selects his own special branch of general education, leaving the other branches to his fellow members of society in the expectation that they will share the benefits of *their* education with *him*, just as he will share the benefits of *his* with *them*." [*ibid.*] The basic idea behind the division of classes is that a specific class can be fully dedicated to the cultivation of each specific talent that characterize men. These talents can be co-ordinated with each man's natural needs and it is the satisfaction of these needs that is the task of each specific class. Of course, there is a hierarchy of these needs modeled on the scale of goods that determines the formation of the life-plan of an individual person; as we insisted in the previous section, vital needs may be sacrificed – and it can even become a duty to sacrifice them – for higher needs, such as intellectual goods. According to Fichte, a perfect society is one where *all* needs are cared for, and where they all are developed and satisfied *equally*. Society is perfect *qua* society when it is so organized that "it would necessarily have to *approximate* more and more closely to its goal." [*ibid.*] How is the society to reach its goal and how does the class of scholars in particular contribute towards the task of reaching this goal? Fichte distinguishes three necessary presuppositions for seeing to the equal development of all of man's talents: (1) *a scientific knowledge of*

[32] Italics added.

all of his drives and needs, a complete survey of his entire nature ("What is man?"), (2) *the scientific knowledge of how to develop and satisfy them* ("What must man do?") and (3) *knowledge of the particular cultural level of one's society at a particular time and the events of former ages that have led to this particular stage of development* ("Whence did man come from?"). These three types of knowledge constitute what is called "learning": "the person who dedicates his life to the acquisition of such knowledge is called a 'scholar'." [*ibid.*, 172] It is obvious that every individual scholar could not – and should not – take as his occupation the entire field of human knowledge, and *a fortiori* not in all three of the aspects mentioned above. The division of labour is a valid principle also within the confines of science and individual scholars can stake out for themselves individual portions of the domain of knowledge, but then "in his own area each person should cultivate all three: *philosophical* and *philosophical-historical*, as well as purely historical knowledge." [*ibid.*][33] Fichte then proceeds to the high point of his lecture where the true vocation of the scholarly class is indicated:

> [T]he study of a properly grounded philosophy does not make it superfluous to acquire empirical knowledge – not, at least, if such knowledge is thorough. On the contrary, such a philosophy demonstrates in the most convincing manner the indispensability of empirical knowledge. We have already shown that the purpose of all human knowledge is to see to the equal, continuous, and progressive development of all human talents. It follows from this that the true vocation of the scholarly class is the *supervision of the actual progress of the human race in general and the unceasing promotion of this progress.* [*ibid.*]

This passage is remarkable in that it is completely in consonance with the Enlightenment idea of 'social engineering' and the ideal of the progress of mankind that is facilitated by a thorough knowledge of the sciences. Contrary to the unfortunately common misinformed readings of Fichte, we can here see an explicit commitment to the significance of the empirical. It is no coincidence that the Fichtean ideal is deeply ingrained with the idea that "[t]he whole progress of the human race depends directly upon the progress of science". Indeed, the Kantian roots of Fichte's systematic philosophy are well-known and the political circumstances of the society in which he developed his ideas were highly conducive to the kind of *utopian* elements that were characteristic of the Enlightenment political philosophy.

From the modern point of view the requirement that the scholar is supposed to "supervise and promote the progress of the other classes" may seem a lofty and unrealistic idea. There is, however, something intrinsically valid in the idea that must be brought to light and considered seriously: the scholar is in a unique position because he is, so to speak, "especially destined for society". As Fichte remarks, "More than any other class, his class, insofar as he is a scholar, properly exists only through and for society. Accordingly, it is his particular duty to cultivate to the highest degree within himself the social talents of *receptivity* and the *art of communication*." [*ibid.*, 173] Here we encounter what, once

[33]Italics added. The three types of knowledge correspond, naturally, to the three presuppositions above; (1) philosophical, (2) philosophical-historical and (3) historical, respectively.

again, could be accommodated within a system of virtues. Although these talents are not strictly speaking epistemic virtues, they can be said to be necessary conditions for exercising such virtues. The requirement of communicative rationality that is conveyed by Fichte's remarks, is in excellent agreement with the narrative of the tradition of epistemic virtues provided by Daston and Galison. Especially noteworthy are Fichte's comments on the virtue of (intellectual) tolerance that he sets as a basic *desideratum* for every scholar; a scholar "should be familiar with his scientific predecessors. And this familiarity cannot have been produced merely by rational reflection, but has to have been learned through oral or written instruction. By constantly learning something new he should *preserve his receptivity and try to guard against that total lack of openness to foreign opinions and ways of thinking* which one often encounters, occasionally even among excellent and independent thinkers. For no one is so well instructed that he could not always learn something new and occasionally something essential." [*ibid.*, 173–174][34] In addition to the communicative skills, a scholar should develop a sense for what is true.[35] In as much as all men share this fundamental capacity at least to some degree, it is not sufficient in itself. Indeed, "[i]t has to be developed, scrutinized and purified, and this is precisely the scholar's task." [*ibid.*] But most remarkable of all is Fichte's conception of the scholar as "the *teacher* of the human race". Apart from making men generally acquainted with their needs and means for satisfying them, and directing their attention to the needs which confront them as well as the specific means for achieving each purpose, the scholar also sees the direction in which human race must proceed. In his role as the educator of mankind, "the scholar is subject to the ethical law, which commands harmony with oneself." [*ibid.*] In consonance with the final aim of every individual person, the scholar must strive to work for the "ethical improvement of the whole person", always keeping this ideal in his view, taking it as a guideline in everything that he does in the society, because "no one who is not himself a good man can work successfully for ethical improvement". Fichte clearly perceives the truth that the importance of moral examples goes beyond the elaboration of ethical theories, and this can be generalized. Personal role models do not provide the justification for the validity of moral judgements but they are important precisely because they *motivate* people to act morally, and often in a significantly deeper way than any (even most complete) ethical grounding ever could. As Fichte remarks, "[e]veryone who lives in society owes it to society to set a good example, because the power of example originates only through our life in society." [*ibid.*][36] We have already seen some implications of these ideas to the broader panorama of the ethics of science. Fichte's lectures can be conceived as a courageous elaboration of the theme of "scientific self", the notion whose importance we have stressed above. Nevertheless, Fichte's ideas on the scholar's vocation go significantly beyond the cornucopia of different selves that Daston and Galison associated with the different techniques of the self which have their origin in the various epistemic practices of science. These views do not contradict each other, however. The Fichtean idea that the penultimate duty of a scholar is to strive to be the "ethically best

[34]Italics added.

[35]Fichte later developed further the notion of *Wahrheitsgefühl*, or the "feeling" or "sense of truth" in his (unpublished) lectures.

[36]In this connection Fichte quotes the words of the founder of Christianity, which apply quite aptly also to the scholars: "Ye are the salt of the earth, but if the salt has lost its savor wherewith shall it be salted?" [Matt. 5:13]

man of his time" is nothing but an expression of the formal function of self-attribution, i.e., it posits the normative requirement of self-determination. The specific characteristic of *I* is that it can enjoy a plurality of selves, hence also different "scientific selves". The Fichtean idea is understood correctly if it is seen as operating on a meta-level where one confronts one's *descriptive self-image* with a *normative self-image*. Vittorio Hösle's remarks about these notions are especially relevant here:

> The *I* not only entertains assumptions regarding what it is, but also regarding what it ought to become. The *I* can only accept its self if the descriptive self-image corresponds to the normative self-image; however, this harmony is justified only if both self-images are also truly rational, that is, if the *I* really knows its self and if its normative self-image is appropriate. [...] on one hand, the *I* must recognize some values that are general or at least transcend it; on the other hand, the recognition of general values in no way resolves the question of what normative self-image I should have of myself. I may well recognize that I should become a useful member of the society, but what line of work I should go into remains undetermined. For that, I need information about myself [...] [Hösle (2004), 247–248]

Completely in agreement with the Fichtean point of view (and the articulation of the meaning of the process of self-attribution given by Hösle), Daston and Galison stress that "the mastery of scientific practices is inevitably linked to self-mastery, the assiduous cultivation of a certain kind of self". [Daston & Galison (2007), 40] Furthermore, they explicitly bring forth the relationship between epistemic practices and traditional, philosophical and religious forms of self-cultivation: "objectivity is to epistemology what extreme asceticism is to morality. [...] The demands it makes on the knower outstrip even the most strenuous forms of self-cultivation, to the brink of self-destruction. [...] It is a sacrifice." [*ibid.*, 374]. This passage, along with the references to Hadot's work on the philosophy of antiquity, are reminiscent of the words – quoted by Hadot himself – of the Christian monk Dorotheos of Gaza (6^{th} century CE): "He who has no will of his own always does what he wishes. For since he has no will of his own, everything that happens satisfies him. He finds himself doing as he wills all the time, for he does not want things to be as he wills them, but he wills that they be just as they are." The requirements posited for the scholar, remarkably similar in spirit to the moral maxims of contemplative and religious practise, are thus extremely stringent. But then, they are posited as ideals that are *regulative* rather than *coercive*; they are posited in order to direct the practise and moral conduct of scientists in ways that, to paraphrase Daston and Galison, facilitate the internalization and enforcement of norms that are critical in securing knowledge. It is a truism hardly meriting an explicit statement that ideals cannot be actualized in the real world. Still, the fact remains that for the improvement of the human condition it is of utmost importance to realize that reality must be judged in accordance with ideals and modified by men who feel themselves capable of doing so. Fundamentally, the problem of the realization of ideal validities can be traced back to the distinction between *Is* and *Ought* which is so crucial for the elaboration of the many questions of theoretical and practical ethics. These dualisms, although in many ways different, are connected and their dynamic interaction in the course of the historical development of science is one of the most crucial elements of any philosophically respectable account of the progress of science in the twentieth century.

In this section we have seen how the Fichtean conception of the scholar's vocation dovetails with the tradition of epistemic virtues in the history of science. We have also seen how, in surprising agreement with the Enlightenment ideals that underpin Carnap's work, Fichte stressed the significance of the virtue of tolerance for the scholar. Since the notion of tolerance constitutes an essential part of the system of virtues, and since it also figures prominently in the philosophical thought of Carnap, it is appropriate to devote a separate section to the Principle of Tolerance and its position within the history of morality. This is the subject of the following section.

2.4 Universalism and the Principle of Tolerance

"Reason is the special embodiment in us of the disciplined counter-agency which saves the world" — A. N. Whitehead, *Function and Reason*, Ch. *I*

It is well known that the notion of tolerance, understood as a philosophical concept pertaining to morals and politics, has its intellectual origins in the philosophical debates of the seventeenth century. It first surfaces in the philosophical works of Locke and Spinoza. These two thinkers differ considerably according to their evaluation of the systematic importance and general significance of the concept. The tensions between them exemplify in an interesting way the fundamental cleavage between the conservative and radical forms of Enlightenment. The historical significance of the principle is reflected in the manner in which it paves the way for the modern concept of individuality. In Spinoza's thinking tolerance is promoted to the status of a regulative principle that constitutes an important part of the framework in which power structures and forms of political action are legitimated. Furthermore, its scope comprises not only the domain of morals and politics, but also the domain of theoretical thought taken in its most general sense of reflecting formal-universal laws. Indeed, the "Principle of Tolerance" is most naturally conceived as a vehicle for increasing intellectual freedom within the society, an idea that is inextricably entwined with Spinoza's political ideal of promoting democracy as the rationally legitimated form of constitution.

As Spinoza's philosophy may be argued to comprise the intellectual backbone of the movement of Radical Enlightenment – as Jonathan Israel has attempted to demonstrate [Israel (2001), *passim.*] — his formulation of the Principle of Tolerance constitutes, then, a natural point of departure for any historical survey attempting to shed light on the different philosophical concepts centered around the Enlightenment concept of intellectual virtues. In this section I will primarily concentrate on the notion of tolerance and its central philosophical importance in the thought of Spinoza. The issues he takes into consideration are completely novel, as we will see. Before 1670s, the term "tolerance" had a much more restricted sense. In Latin it was associated with the virtue of fortitude.

It had both a literal and a moral meaning. In a moral sense it meant steadfastness. In the most literal sense it referred to physical endurance. Indeed, taken in its physical sense, to tolerate something quite literally is to bear something. And analogously, in the moral case, tolerance is steadfastness or endurance in the face of some evil or hardship. In the seventeenth century the notion of tolerance becomes markedly an object of political and religious discourse, the shift being largely bred by the tumultuous climate of religious conflicts brought about by the Reformation. Political battles abounded, not to mention wars and revelries. A significant occasion was the twenty-fourth of May in 1689 when the British Parliament passed the *Toleration Act* that allowed dissenting Protestants freedom of worship. The parting shot for the philosophical debate on tolerance was the publication of the first volume of John Locke's *Epistula de Tolerantia* in 1689. The interpretation of tolerance which it propounded was in a glaring opposition to the radical conception of Spinoza. For a long time, though, Spinoza's sustained arguments for adopting a broad conception of the content of "tolerance" and for propounding the significance of toleration in his *Theologico-Political Treatise* (published in 1670) were counterproductive. As W.N.A. Klever puts it, the work "neither helped to reduce the influence of theological prejudices among philosophically-minded readers nor furthered the freedom required for the enlightened citizen. In fact, the *Theologico-Political Treatise* did not at all prepare the way for the publication and reception of Spinoza's overall philosophy. Its publication, on the contrary, aggravated the situation by unchaining a series of devastating refutations and defamations." [Klever (1996) [Garrett (1996), 39]] Thus, initially the constructive influence of Spinoza's ideas was limited to a small circle of intellectual radicals. The situation did not improve very much in the eighteenth century which was dominated largely by a moderate version of Enlightenment. Ironically, it is perhaps mostly due to the vindication of Spinoza by Hegel, whose philosophy of objective idealism is the epitome of German Romantic thought, that Spinoza's philosophical contributions were more and more appreciated in the nineteenth century. Indeed, as is familiar from popular accounts of history of philosophy, Hegel famously articulated the choice imposed upon any philosopher: either Spinozism or no philosophy at all. However, in spite of Spinoza's *Ehrenrettung* by Hegel, the road of Spinozism in Western thought was a rocky one, consisting of many twists and turns; from the historian's viewpoint the reception history of Spinozism is complex, having a myriad of subplots, and it is not easy to form a coherent narrative of its growing influence in Europe.[37] I cannot, then, if only because of limitations of space, delve into this particular topic in any detail.

A somewhat surprisingly neglected issue in the history of ideas is the constitutive function of the notion of tolerance within the domain of formal-universal thought from the seventeenth century on (we mean here its significance for theoretical and speculative thought in general, and for the exact sciences in particular). The extension of the scope of the Principle of Tolerance from the exclusively ethical and political questions, primarily pertaining to social welfare and equality, to the epistemic questions of science, concerning the receptivity to novel and unfamiliar modes of thinking as well as evaluation of different technical language systems, e.g. the evaluation of the Newtonian and Leibnizian

[37] For a comprehensive presentation of the history of Spinozism in the Enlightenment, see Jonathan Israel's splendid book *Enlightenment Contested* [Israel (2006)].

systems of differential and integral calculus, marks an important shift in the development of epistemic virtues in European thought. It is precisely the fusion of two qualitatively different kinds of discourses, the ethical and the epistemic, that makes Spinoza's philosophy a paradigmatic exemplar of Radical Enlightenment thought. However, Spinoza never made explicit the full ramifications of his notion of tolerance, and it has largely remained a task for later generations to develop his ideas further. In this respect the notion of tolerance – and one of the goals of this work is to adduce this more clearly and render it plausible – constitutes one of the regrettably neglected concepts in the history of ideas with regard to its philosophical significance. The radical shift in the *meaning* of this notion is epitomized by the historical dialectic between the ideas of Locke and Spinoza, as noted above. However, due to the influence of moderate interpretations of Enlightenment ("moderate mainstream Enlightenment", as Jonathan Israel likes to call it) and movements of Counter-Enlightenment, the central issues, including the essential interplay between ethical and epistemological aspects of tolerance, as well as the more general interplay between epistemic and moral virtues, failed to come across in modern philosophy up until the turn of the twentieth-century (or, depending on which philosophers one concentrates on, even the twenty-first century). Until quite recently, the neglect of the study of the history of intellectual virtues in general has also contributed towards this desultory state of affairs. Fortunately, recent work by some scholars (notably Lorraine Daston and Peter Galison) in the history of science has, once again, opened up the subject for more detailed and systematic discussion. The tradition of intellectual virtues is closely linked with the modern history of science, the emancipation of the *scientific self* and the ethics of knowledge, as we have seen above. The Principle of Tolerance is an inextricable part of these general developments.

2.4.1 Spinoza on tolerance

The following is a famous passage from Spinoza's *Theologico-Political Treatise*:

> How much better would it be to restrain popular anger and fury, instead of passing useless laws, which can only be broken by those who love virtue and the liberal arts, thus paring down the state till it is too small to harbour men of talent. What greater misfortune for a state can be conceived than that honourable men should be sent like criminals into exile, because they hold diverse opinions which they cannot disguise? What, I say, can be more hurtful than that men that have committed no crime or wickedness should, simply because they are *enlightened*, be treated as enemies and put to death, and that the scaffold, the terror of evil-doers, should become the arena where the highest examples of *tolerance* and virtue are displayed to the people with all the marks of ignominy that authority can devise?
>
> He that knows himself to be upright does not fear the death of a criminal, and shrinks from no punishment; his mind is not wrung with remorse for any disgraceful deed: he holds that death in good cause is no punishment, but an

> honour, and that death for freedom is glory.
> What purpose then is served by the death of such men, what example is proclaimed? the cause for which they die is unknown to the idle and the foolish, hateful to the turbulent, loved by the upright. The only lesson we can draw from such scenes is to flatter the persecutor, or else to imitate the victim.
> If formal assent is not to be esteemed above conviction, and if governments are to retain a firm hold of authority and not be compelled to yield to agitators, it is imperative that freedom of judgement should be granted, so that men may live together in harmony, however diverse, or even openly contradictory their opinions may be. We cannot doubt that such is the best system of government and open to the fewest objections, since it is the one most in harmony with human nature. In a democracy (the most natural form of government [...]) everyone submits to the control of the authority over his actions, but not over his judgement and reason; that is, seeing that all cannot think alike, the voice of the majority has the force of law, subject to repeal if circumstances bring about a change of opinion. In proportion as the power of free judgement is withheld we depart from the natural condition of mankind, and consequently the government becomes more tyrannical. [Spinoza: *TTP* xx [PT (1951), 263–264]][38]

Spinoza's exhortation, mainly aimed at those in possession of political power and capable of passing decrees, to keep constantly in mind men's irrevocable right to freedom of judgement, is one of the very first expressions of the content of the notion of tolerance that is the hallmark of Radical Enlightenment. The exhortation is a serious one: the personal situation in which Spinoza found himself in Amsterdam in 1670, notwithstanding the fact that he lived in one of the most liberal of the European states, was awkward. Until the late 1660s Spinoza had seen Amsterdam as a quite positive example of arranging public life: "In this most flourishing republic and excellent town people of all nations and sects live together with highest unanimity." [*TTP* xx.40] While Spinoza belonged to a group of political refugees, the minority of Portugese Jews (Sephardic Jews), he had not had any notably bad experiences with state authorities and their judicial system. This situation was not to last, however:

> As Spinoza was writing his treatise, the situation worsened because of a serious economic malaise and the political isolation of the Dutch Republic. Intolerance was aggravated also, and came very close to Spinoza himself. To the circle of his friends and followers belonged a certain Adriaan Koerbagh, who had studied medicine and law in Utrecht and Leiden. He had (with his brother Johan) become persuaded by Spinoza's naturalism, and was also acquainted with Franciscus van den Enden. This man, only two years younger than Spinoza, started to spread all the essentials of the Spinozistic theory from 1665 onward, and published them in 1668, in plain Dutch. His main work, *Een ligt*, was on many pages more open about Spinoza's esoteric doctrine than the

[38] Italics added.

> *Theological-Political Treatise*. God is defined as "the essence of all modes of existence, consisting of infinite attributes, of which each one is infinite in its kind." The work as a whole may be considered as a parallel to the *Theological-Political Treatise*, with chapters on Essence (God is consequently called "*Wesen*"!), the Savior (Jesus), the Holy Spirit (reason), good and evil, religion, the Bible, heretics, heaven, miracles, and so on. Nowhere in the book is Spinoza's name mentioned, but his doctrine is elaborated on many pages. The author was, however, much less prudent than Spinoza himself, and launched strong attacks on the preachers and theologians. When he was arrested, he confessed at his trial to his relations with Spinoza and Van den Enden. He was sentenced, in the free town of Amsterdam, to ten years in the house of correction (*truchthuis*), ten years exile afterward, and a fine of 6,000 guildens. He was thrown into the *Rasphuis*, a prison with very bad circumstances, in which he died a year later (in 1669). [Klever (1996), 38–39]

Instigated by this incidence, which must have made a deep impression on him, Spinoza, who was occupied writing his *Theological-Political Treatise*, included in the book a passage, probably referring to the fate of his friend: "Following this example of the Pharisees [having accused The Sadducees of impiety] the vilest hypocrites, agitated by the same frenzy (*rabie agitati*) which they call zeal by divine right, have always persecuted men distinguished by their honesty and their virtue and therefore envied by the mob; they do this by publicly despising their opinions and inflaming the anger of the furious multitude against them." [*TTP* xviii.24] Against this background it is clear that Spinoza's political writings, most prominently the *Theological-Political Treatise*, were not mere theoretical exercises, i.e., mere philosophical conclusions derivable from his system. Nay, they were very much born through practical considerations, stemming largely from his personal experiences. In this sense, the distinction between *moral virtue* and *ethical virtue* is a relevant one; Spinoza, deeply shocked by the train of events, exemplified not only genuine sympathy for the sufferings experienced by Koerbagh, but raised the moral feelings of sympathy to the communicative form of ethical insights that he included in his treatise. On a more general level, it can be said that the significance of moral feelings consists precisely in their unique capacity to induce mental qualities of intensity that the purely theoretical thought is vainly longing for. In any case, genuine sympathy can be said to be an exceptionally complex intentional act, because it operates, so to speak, on a meta-level; in contradistinction to other (moral) feelings it is directed to the pain *simpliciter* experienced by another person, and not to the *cause* of that pain. Its operation is based on the *I*'s capacity for abstraction which enables it to consider the other selves as being of equal value with itself. Therefore, the transcendental dimension of morality is clearly operative in instances of genuine sympathy.

What else can be distillated from the passage quoted above? The lines concerning the necessity of granting freedom of judgement to every citizen of a state, accompanied with the passing mention about tolerance, hint at yet another important dimension of Spinoza's political thought, viz., the necessity of submitting to the control of the authority in matters of action. This constitutes a notable tension in Spinoza's account. *Prima facie*, it would seem that if "the voice of the majority has the force of law", there is not much room for the radical liberalism that is advocated by Spinoza in the very same passage. The appearance of a contradiction evaporates, however, as soon as one recognizes the background framework influencing Spinoza's thought. As is well known, Spinoza derived much of the philosophical content of his system from Radical Cartesianism. An important example is the Radical Cartesian doctrine of self-interest about which Spinoza provided a metaphysical understanding. According to Spinoza, all things, including humans, strive to persevere in their existence. In order to survive, humans must come together and form a society. Without mutual aid, people would not have the time and skill to support and preserve themselves to the greatest possible extent. [*TTP* v.73/13] And so people come together, giving up the rights that they have in the state of nature, and agree to obey the laws of the state. If people only desired what is prescribed by true reason, the society would need no laws:

> Nothing would be required to teach men true moral doctrine, and they would then act to their true advantage of their own accord, wholeheartedly and freely. But human nature is far differently constituted. All men do, indeed, seek their own advantage, but by no means from the dictates of sound reason. For the most part the objectives they seek and judge to be beneficial are determined only by fleshy desire, and they are carried away by their emotions, which take no account of the future or of other considerations. [*TTP* v.74/31]

In this sense society needs governance and coercion; it needs laws to control and restrain the people's lusts and urges [*ibid.*] The laws must be set up in a way that, whether they will it or not, people act in the interests of the common welfare. The state must not depend on human virtue (presumably short of the strictures of true reason), but rather, cause human virtue through necessity. Spinoza's conception about the constitution as "the soul of the state" is made more comprehensible by the following passage:

> It is not, I repeat, the purpose of the state to transform men from rational beings into beasts or puppets, but rather to enable them to develop their mental and physical faculties in safety, to use their reason without restraint and to refrain from the strife and vicious mutual abuse that are prompted by hatred, anger, or deceit. Thus the purpose of the state is, in reality, freedom. [*TTP* xx.241/3]

Whereas men are obligated to frame their decisions for action according to the (positive) laws and orders devised by the local legislative bodies and authorized by the local government, they are guided in their expectations of future by the regulative principles of thought; principles that are articulated and embodied in the current practices of science (including the generally accepted conventions of presentation, communication and criticism of ideas). The basis for evaluation of the adequacy of these two spheres, the sphere of social practice and the sphere of knowledge, resides in the force of the freedom of judgement. This freedom, in turn, depends on the requirement of *rationality* that comprises the most general conditions upon thought and action. While it is not plausible to assume that the judicial and legislative practices could account for the more general and universal principles of rationality and ideal validity in all of their aspects, it is mandatory – if just to abstain from anarchy – to conform to the prevailing laws and customs. This compromise is accepted with the hope of gradually improving and revising the legislative practices so as to make them more compatible with the strictures of Reason, i.e. the domain of pure validities.

Spinoza's conception of rationality strikes us today as surprisingly modern. Spinoza took reason to be a cognitive and effectively inferential process that was applied in order to derive adequate knowledge from some other piece of adequate knowledge. Clearly, such a conception, although superficially stated in terms similar to deductive reasoning, comprised a large variety of inferential processes, clearly outnumbering the 'merely' deductive ones. Included among the tasks of (practical) reason are, according to Spinoza, guidance of action (*ex ducto rationis*), counsel offering (*consilium rationis*), providing precepts (*præceptum rationis*), rules (*præscriptum rationis*), and 'dictates' (*dictamen rationis*). One is immediately struck by the resemblance of this scheme to ideas that comprise the core of modern cognitive science: in the first place, Spinoza distinguishes between numerous different *functions* of reason (the different cognitive capacities of man), and in the second place, he emphasizes the *fallibility* of human reason (the restrictions of information processing in the human brain). Of course, there are many more aspects to the intricate relations between Spinoza's thought and modern cognitive psychology, but these are the most important.[39] The essence of the doctrine of reason in Spinoza is that in order for it to act within the sphere of morality, it must have motivational force. How is Spinoza able to explain the sources of this motivational force? The sources relate to the basis of rational life which Spinoza thinks is the ultimate characteristic of humanity in its highest form. But man alone could not conduct his life according to the demands of rationality if there was not a motivational force that sprang from a much more powerful source:

> But human power is extremely limited, and is infinitely surpassed by the power of external causes; we have not, therefore, an absolute power of shaping to our use those things which are without us. Nevertheless, we shall bear with an equal mind all that happens to us in contravention to the claims of our own advantage, so long as we are conscious that, we have done our duty, and that the power which we possess is not sufficient to enable to protect ourselves completely; remembering that we are a part of universal nature, and that we follow her order. If we have a clear and distinct understanding of this, that part of our nature which is defined by intelligence, in other words the better part of ourselves, will assuredly acquiesce in what befalls us, and in such acquiescence will endeavor to persist. For, in so far as we are intelligent beings, we cannot desire anything save to that which is true: wherefore, in

[39]One interesting contemporary attempt to link Spinoza's philosophy to modern cognitive science is Antonio Damasio's *Looking for Spinoza: Joy, Sorrow and the Feeling Brain* [Damasio (2003)], where the Spinozist account of *affects* is linked with modern neurobiological research on emotions.

> so far as we have a right understanding of these things, the endeavor of the better part of ourselves is in harmony with the order of nature as a whole. [*E* iv.p.xxxii]

Another important dimension of the Principle of Tolerance advocated by Spinoza, inextricably entwined with the one sorted out above, relates to historiography and interpretation of texts. The views of Spinoza on this topic had their origin in the assessment of the function of hermeneutics in Spinoza's overall philosophy. Indeed, the interpretation of various religious and historical texts constituted according to Spinoza a necessary requirement in an epistemic endeavor to acquire better and more reliable knowledge about nature and man. The hermeneutic approach constituted an important stage in the middle of an individual's emancipation from false beliefs and rigid authorities. As Spinoza put it:

> If we read a book which contains incredible or impossible narratives, or is written in a very obscure style, and if we know nothing of its author, nor of the time or occasion of its being written, we shall vainly endeavour to gain any certain knowledge of its true meaning. For being in ignorance of these points we cannot possibly know the aim or intended aim of the author; if we are fully informed, we so order our thoughts as not to be in any way prejudiced either in ascribing to the author or him for whom the author wrote either more or less than his meaning, and we only take into consideration what the author may have had in his mind, or what the time and occasion demanded ... It often happens that in different books we read histories in themselves similar, but which we judge very differently, according to the opinions we have formed of the authors ... Thus it is evidently necessary to know something of the authors of writings which are obscure or unintelligible, if we would interpret their meaning; and for the same reason, in order to choose the proper reading from among a great variety, we ought to have information as to the versions in which the differences are found, and as to the possibility of other readings having been discovered by persons of greater authority. [*TTP* vii [PT (1951), 111-112]]

It is tempting to interpret this passage as an early modern formulation of the *principle of charity* the central significance of which Donald Davidson has emphasized for the theory of argumentation. Occasionally referring to it also as the *principle of rational accommodation*, he has summarized its content as follows: "We make maximum sense of the words and thoughts of others when we interpret them in a way that optimises agreement." [Davidson (1974)][40] It is fascinating to see that this essential aspect of modern (linguistic) pragmatics is clearly anticipated in Spinoza's writings concerning hermeneutics.

In order to render the picture of Spinoza's notion of tolerance complete, it is essential to give a summary view of the differences of the interpretations of it provided by him

[40] Davidson, Donald (1984) [1974]: "Ch. 13: On the Very Idea of a Conceptual Scheme", *Inquiries into Truth and Interpretation* (Oxford: Clarendon Press).

and Locke. The two men differed radically on their opinions about the meaning and implications of tolerance. Giving an account of these differences is illuminating, for it illustrates the two entirely different stances that these gentlemen, the most illustrious harbingers of Enlightenment, took in regard to the notion that was to have fundamental repercussion in philosophy in the next few hundred years. I will first give a general outline of Locke's moral philosophy.

Essay, Book II, Chapter XXVIII, §§7–14

In the beginning of the section §7, Locke represents a division of laws which he thinks people generally refer their actions to. This division amounts to the following:

1. The *divine* law.
2. The *civil* law.
3. The law of *opinion* or *reputation*.

The first type of law is the standard for 'sins, or duties', the second for 'criminal, or innocent' actions, the third for 'virtues and vices'. [Leibniz, *New Essays*, Book II, Chapter xxviii, §7] There are some very interesting issues that pertain to the classification of laws in this manner. To start with the most perspicuous ones, we may raise the points that Leibniz already made in his famous commentary to Locke's *Essay*. [*ibid.*] Firstly, a semantical distinction with respect to sins and duties amounts really to a distinction between dispositions (which Leibniz termed *habitudes**[41]) and actions. Secondly, issues pertaining to virtue and vice do not usually depend on general opinion. On the contrary, within the context of Christian ethics in which the discussions between Locke and Leibniz are embedded, the concepts of virtue and vice are predominantly defined by theology, i.e. overtly by the *divine* law, and covertly by the ecclesiastical authorities. They are thus to be understood as strictly deontological notions. This is, understandably, congruent with Leibniz's theism. Thirdly, there are two sorts of divine law: natural and positive.[42] Civil law is positive. Fourthly, the 'law of reputation' cannot properly be called a law unless it is included in the natural law — as one might speak of the 'law of health', 'the law of business', in areas where one's actions naturally bring various goods and evils, such as the approval of others, health, monetary gain.

[41]The commentary by Peter Remnant and Jonathan Bennett (in NE) explicates this concept in the following way: "This can mean 'habit', 'custom', 'character-trait', 'tendency', 'general disposition'. In most occurrences 'disposition' seems right; but where '*disposition*' is also present we render '*habitude*' by 'tendency' instead. Sometimes [...] '*habitude*' is used to contrast not dispositional with actual, but rather general with particular, within the realm of actual behavior. In these contexts we use 'habit', or 'practice', or, once, 'regularity of conduct'."

[42]Hobbes gives an account of this distinction: "[...] *Natural* are those which have been laws form all eternity, and are called not only *natural*, but *moral* laws ... *Positive* are those which ... have been made laws by the will of those that have had the sovereign power over others ... *Divine positive laws* ... are ... the commandments of God, not from all eternity, nor universally addressed to all men, but only to certain people, or to certain persons". [*Leviathan* II.26]

In the Draft B for the *Essay* Locke gives a very interesting account of the ways in which we acquire moral ideas, i.e. ideas of virtues & vices:

> [. . .] But these Ideas of virtues & vices being of transient actions noe where permanent but only the Ideas conceived in our minds to examine & denominate our actions by we cannot by the immediate information of our senses conversant about reall things get a notion of them & therefor must attein them one of these 2 following way(s)
>
> §**157** 1° Either by the common consent & usage of the country & those men whose language we speake. For if there were noe law, noe punishment noe obligation humane or divine yet there must & would be in the Societys of men notions of virtues & vices Justice Temperance & Fortitude &c consisting in certain collection of simple Ideas without which notions all those words which expresse morall things would in all languages be perfect jargon & insignificant. But all the knowledge of particular virtues & vices which a man atteined to this way would amount to noe more then taking the definitions or significations of the words of any language either from the men|skilled

According to Locke's friend James Tyrrell, the idea for the *Essay* originated in a discussion about 'the principles of morality and revealed religion'. Contrary to popular accounts of Locke, his philosophy is not to be described as plainly empiricist (whatever that means). By contrast, if rationalism is taken to imply that knowledge *par excellence* forms a deductive system, then Locke is a rationalist as much as any other seventeenth-century philosopher. Indeed, paradoxically it is Locke, not Leibniz, who maintains that all genuine knowledge is of *a priori* truths.[43] Locke's conception of moral relations amounts to this:

> For certainty being but the perception of the agreement or disagreement of our ideas and demonstration nothing but the perception of such agreement by the intervention of other ideas or mediums; our moral ideas, as well as mathematical, being archetypes themselves and so adequate and complete ideas; all the agreement or disagreement which we shall find in them will produce real knowledge, as well as in mathematical figure. [*Essay*, IV, iv, 7]

I do not intend to investigate the detailed arguments that Leibniz presented in his *Noveaux Essais sur l'entendement humain*, the extensive commentary to Locke's *Essay*, as a criticism of Locke's position. Suffice it to say that the weakness in the theoretical basis of Locke's moral philosophy was destined to radiate into issues concerning the applications of that philosophy. It is therefore no wonder that Locke's concept of tolerance became, ultimately, equally an issue of controversy. Locke's *Epistula de Tolerantia* (*Letter*

[43] According to Locke, knowledge about the external world is an 'assurance that *deserves the name of Knowledge*'. [*Essay*, IV, xi, 3]

Concerning Toleration) was a defense of conventional, ecclesiastical views, which could not accommodate the sort of radical interpretation of tolerance that compromised the overall legitimating power of one privileged religious creed. Spinoza's interpretation, on the contrary, was of an all-encompassing and universal variety; strictly ecclesiastic issues would no longer be the primary concern of tolerant, rational discourse. Let me now briefly describe the contents of the controversy.

2.4.2 Locke, Spinoza and the philosophical debate on toleration

Locke wrote his *Letter* in November and December 1685 during the period of his most intensive involvement with his Arminian friends — as well as with the Jewish controversialist Isaac Orobio de Castro — and the time of the international *furore* surrounding the Revocation of the Edict of Nantes and the persecution of the Huguenots in France. Locke's notion of toleration can be characterized briefly as being primarily concerned with freedom of worship, of religious practice, as an extension of freedom of conscience, rather than with freedom of thought, speech and of the press. This is the main point of disagreement with Spinoza. Moreover, Locke's conception of toleration includes a curious mixture of biased attitude and theoretical thought. Indeed, Locke is grudging about extending toleration to certain groups and positively hostile to toleration of certain others. [Redwood (1976), 83] In particular, there can be discerned three limitations to Locke's notion of toleration that are noteworthy, as has been shown by Jonathan Israel:

> Firstly, because his toleration is essentially what one scholar has called a 'privilege' and 'immunity' exempting religious dissenters from the church otherwise prescribed and generally obligatory for the whole people — the state church established by crown and Parliament — toleration can only formally and expressly be granted to categories of the population possessing an organized, publicly acknowledged and constituted form of worship for which immunity can be claimed, Protestant dissenters in the first instance but potentially at least also Catholics, Jews and Muslims. Persons subscribing to no recognized church or sect, by contrast, be they agnostics, sceptics, deists or 'Indifferenti' while not specifically excluded are left in a vague limbo without any precise status or acknowledged freedom. [...] Secondly, there is Locke's well known reluctance to accord toleration to Catholics. Strictly speaking, being a defined confession and organized church, there should be no difficulty about extending toleration to Catholics, and indeed Episcopius and, somewhat later, also Uyttenbogaert do so explicitly. But on this point, adhering to the view he had held for many years, Locke is at the very least grudging if not positively intolerant. [...] A third important respect in which Locke's toleration is limited is his emphatic and absolute exclusion of atheists. Since they do not believe in a providential God and belong to no recognized form of worship, and are not seeking to save their souls, by definition they are not entitled to toleration. [Israel (1999), 9–11]

This passage makes evident the contrast between Spinoza's and Lockes's theories of toleration. First of all, Spinoza's account stands out within the intellectual fervent of the early Enlightenment as *not* being built on any theological foundations at all. It is truly a *libertas philosophandi*. For Spinoza, freedom of religion and religious practice are at most a secondary issue. This fact is betrayed by the fact that in the *Tractatus Theologico-Politicus* in which the notion of tolerance is discussed, freedom of exercise of one or another faith is not dealt with at all. Second, in Spinoza, toleration is overridingly about individual freedom and emphatically not the 'freedom' of ecclesiastical structures to confessionalize and claim authority over individuals. As Jonathan Israel has succinctly pointed out, "[t]he most fundamental difference between Spinoza's and Locke's theories of toleration lies precisely in the relegation of freedom of worship, freedom to express religious doctrines, and freedom to organize churches to the periphery of the debate, the emphasis, in other words, is on obviating the formation of powerful ecclesiastical hierarchies and authority. In this respect, Spinoza's concept of toleration forms a part of the wider Dutch republican political thought tradition, a drive to weaken the power of the clergy for the good of society [. . .] " [*ibid.*, 13] I have already alluded to the more detailed contents of Spinoza's notion of tolerance above. What is important to notice here is the social and political context in which that notion emerged and to what a great extent it signified a radical shift in the *habitudes* of thought that prevailed in the early Enlightenment. It is of utmost importance to realize the significance of Spinoza's thinking in this respect for the development of theoretical thought in the sciences. Although it is not immediately clear to his contemporaries that the implications of the notion of tolerance have drastic effects in the sphere of theoretical thought, including natural science, a necessary prerequisite for realizing the true significance of the principle of tolerance is, nevertheless, fulfilled; it is understood that tolerance is not exclusively, not even primarily, an ecclesiastical issue.

The overwhelming importance of the conflict between the views of Locke and Spinoza is therefore captured in the following words of Jonathan Israel:

> The shift from the quest for religious freedom to the quest for philosophical freedom which begins with Spinoza's critically important theory of toleration is, in fact, one of the most characteristic features of the early Enlightenment, the period down to around 1750 when the real business of breaking the confessional structure of *ancien regime* European thought and culture was undertaken. The historical, as distinct from the philosophical, importance of Spinoza's theory of toleration, especially in its contrast with Locke, can hardly be overstated. In the early eighteenth century radical writers in Europe no longer complained about the lack of religious freedom. Suddenly, the brunt of their complaint became the lack of intellectual freedom. [*ibid.*, 19]

Within the confines of Western philosophy and science, reviewing the development of the three hundred years after Spinoza, toleration must be seen mainly as a vehicle for increasing *intellectual* freedom. That the notion was first primarily associated with purely ecclesiastical and political issues renders it particularly hard to grasp how the notion of tolerance became to have the completely different role in twentieth-century philosophy.

The history of intellectual virtues provides the answer to this question, but this history is closely linked with the history of the self, as we have seen. What is Spinoza's contribution to this discussion? The essential content of Spinoza's political writings, as I have tried to demonstrate, is inextricably linked with the program of Radical Enlightenment. The Radical Enlightenment, one of the harbingers of which Spinoza definitively is, and the core ideas of which he very much helped to create in his writings (though of course not under that name), conceived as a package of basic concepts and values, may be summarized in eight cardinal points, as Jonathan Israel has argued [Israel (2006), 866]: (1) adoption of philosophical (mathematical-historical) reason as the only and exclusive criterion of what is true; (2) rejection of all supernatural agency, magic, disembodied spirits, and divine providence; (3) equality of all mankind (racial and sexual); (4) secular 'universalism' in ethics anchored in equality and chiefly stressing equity, justice and charity; (5) comprehensive toleration and freedom of thought based on independent critical thinking; (6) personal liberty of lifestyle and sexual conduct between consenting adults, safeguarding the dignity and freedom of the unmarried and the homosexuals; (7) freedom of expression, political criticism and the press, in the public sphere; (8) democratic republicanism as the most most legitimate form of politics. These are all relevant for he subsequent development of the idea of the modern self, the seeds of which are to be found in Spinoza's thought. They figure prominently in the history of morality and they are pertinent issues when we turn to examine the philosophical thought of the twentieth-century.

Chapter 3

THE SIGNIFICANCE OF *A Priori* AND EMPIRICAL KNOWLEDGE IN ETHICS AND SCIENCE

Throughout the previous chapter we have stressed the importance of the insight, one that already Kant conceived as one of the most important axioms of moral philosophy, that normative judgements (and in an analogous way, value judgements) belong to the class of *a priori* synthetic judgements and thus cannot be either grounded (proved) or falsified by reference to *a posteriori* judgements. The moral maxim of "Act in such a way as to treat humanity, whether in your own person or in that of anyone else, always as an end and never merely as a means" is not affected by considerations about socio-psychological facts or statistical data about human trafficking, for example. However, the negligence which Kant showed towards the empirical in questions of morality is nonetheless detrimental to his system of moral philosophy. Even such a prominent figure as Max Scheler, who built his theory of value judgements on the basis of the notion of substantive *a priori*, underestimated the value of experience and empirical knowledge in ethics. Notwithstanding their important contributions to ethical theory, Kant and Scheler both lack a sense of the importance of experience and empirical judgements in deriving more concrete norms from the system of values. It is clear that on the level of practical action and practical ethics, such knowledge is necessary. The importance of the empirical is grounded on two fundamentally different levels that may be referred to as the level of *subsumption* and the level of *means and ends*. The first concerns, in essence, the question whether a concrete being with respect to which we have to frame our moral decision, is a *person*. This is both historically and systematically relevant question, because many fundamentally moral issues (in history as well as at present) depend upon the answer to it. As Vittorio Hösle correctly observes, "[o]ne reason the answer to the question whether an adult human being is a person is difficult is that the interior dimension of others is not immediately accessible to us. Whether someone has an interior dimension (obviously necessary but not a sufficient condition for being a person) is not so easy to figure out as the question of whether an object has a specific length; *without a theory of the experience of other minds* *this question cannot be fundamentally answered.*" [Hösle (2004), 122][1] Furthermore, his remark on the importance of the process of recognition has significant reverberations for the problem of tolerance as well: "In adult humans the acknowledgment that another is a person is usually connected with the process of recognition (*Anerkennung*) [...] Physical and cultural alterity are barriers to this process of acknowledgment, and overcoming these barriers is a particularly affirmative achievement from a moral point of view." [*ibid.*, 122–123] In summary, then, the importance of the subsumption problem consists in the fact that to be able to answer the question as to what characteristics a (spiritual) being possesses, what features are essential to it in the sense that we will be able to recognize it, and so on, it is necessary to have at our disposal a (non-normative) *theory*. The possibility of such a theory is, of course, a matter of the scientific maturity of the disciplines that are considered to be relevant for answering the question in the first place. The second level concerns the more familiar question about the necessary *means* to ends that are concerned with the protection or realization of goods that men find important. Knowledge about the possible (and ethical) means is naturally empirical, and thus it is quite obvious that a concrete ethics without experiential knowledge is impossible. However, the goals in themselves are not sufficient for evaluating the course of action to be taken; every action has *consequences* the significance of which it is also necessary to assess, and then, this is to be done as extensively as is practically possible. It is a moral duty to foresee as many of the practically relevant consequences of action, but this sets a requirement in the fulfillment of which we quickly encounter limits. The tragic dimension of this issue is aptly described by Hösle:

> It is precisely modern technology's spectacular expansion of the consequences of our action in space and time that makes this duty particularly urgent, and it fundamentally changes the overall ethical situation in comparison with the model proposed by Kant – who, because of his intentionalism, could still assume that common sense was sufficient to understand our duties. The loss of the feeling that one's common sense suffices, together with insight into moral principles, in order to arrive at morally defensible decisions is perhaps the most demoralizing and humiliating trait of the age of technology – a trait that probably best explains the growth of ethical nihilism in an age that is least able to afford it. [*ibid.*, 124]

Notwithstanding the great challenge posed by both of these aspects of the ethically relevant dimensions of empirical knowledge, it is obvious that science has made great progress in developing means to cope with the tremendous complexity of the empirical and social world that directly bears on the two morally relevant methodological levels of the problem of *subsumption* and the problem of *means and ends*. To the extent that the domain of *a priori* knowledge is relevant for *both* science and ethics, it is only natural, and it is certainly wise, to discuss briefly the nature and role of the *a priori* in the history of epistemology. In as much as Carnap's philosophical project is underpinned by the concept of the *a priori* and its relevance for the problem of *means and ends*, it is apt to delve in

[1] Italics added.

some detail into the history of this notion. Let us begin with the notion of analysis that is closely linked with the theme of *a priori* knowledge.

3.1 The Problem of Analysis and *a priori* knowledge

"The long concatenations of simple and easy reasoning which geometricians use in achieving their most difficult demonstrations gave me occasion to imagine that all matters which may enter the human mind were interrelated in the same fashion." — René Descartes

3.1.1 The tradition of analysis

The first systematic uses of analysis as a scientific method we find in ancient Greece. It is well known that the Greek geometers used a method they called *analysis* (ἀνάλυσις) to proceed from a statement of a given theorem to the premisses (and, ultimately, the axioms) from which it could be proved. In these inquiries the method of construction played a significant role. The point of departure was often an auxiliary diagram drawn to include all the given data of the problem together with a sketch of its solution.[2] Analysing these diagrams and the components within them the Greeks were able to extract information out of them pertaining to the conditions and presuppositions for the solution of the problem. This was done in a systematic, step-by-step manner. The first explicit mention of this methodology in a general level can be found in Pappus' works. The famous passage reads:

> The so-called Treasury of Analysis, my dear Hermodorus, is, in short, a special body of doctrines furnished for the use of those who, after going through the usual elements, wish to obtain the power of solving theoretical problems, which are set to them, and for this purpose only is it useful. It is the work of three men, Euclid the author of the *Elements*, Apollonius of Perga, and Aristaeus the Elder, and proceeds by the method of analysis and synthesis. Now analysis is the way from what is sought — as if it were admitted — through its concomitants [the usual translation reads: consequences][3] in order to something admitted in synthesis. For in analysis we suppose that which is sought to be already done, and we inquire from what it results, and again what is the antecedent of the latter, until we on our backward way light upon something already known and being first in order. And we call such a method analysis, as being a solution backwards. In synthesis, on the other hand, we suppose that which was reached last in analysis to be already done, and arranging in their natural order as consequents the former antecedents and linking them

[2] This aspect of the method has been emphasized and adduced in detail in [Hintikka & Remes 1974].
[3] A comment by Hintikka & Remes.

> one with another, we in the end arrive at the construction of the thing sought. And this we call synthesis. Now analysis is of two kinds. One seeks the truth, being called theoretical. The other serves to carry out what was desired to do, and this is called problematical. In the theoretical kind we suppose the thing sought as being and as being as true, and then we pass through its concomitants [consequences][4] in order, as though they were true and existent by hypothesis, to something admitted; then, if that which is admitted be true, the thing sought is true, too, and the proof will be the reverse of analysis. But if we come upon something false to admit, the thing sought will be false, too. In the problematical kind we suppose the desired thing to be known, and then we pass through its concomitants [consequences][5] in order, as though they were true, up to something admitted. If the thing admitted is possible or can be done, that is, if it is what the mathematicians call given, the desired thing will also be possible. The proof will again be the reverse of analysis. But if we come upon something impossible to admit, the problems will also be impossible.
>
> So much of analysis and synthesis. This is the order of the books in the Treasury of Analysis: Euclid's *Data*, one book, Apollonios' *Cutting-off a Ratio*, two books, *Determinate Section*, two books, Euclid's *Porisms*, three books, Apollonios' *Vergings*, two books, his *Plane Loci*, two books, *Conics*, eight books, Aristaeus' *Solidi Loci*, five books, Euclid's *Surface Loci*, two books, Eratosthenes' *On Means*, two books. In all there are thirty-three books . . . [Hintikka & Remes (1974), 8–10]

This conception of analysis that was transmitted from the Greeks to the early modern mathematicians constituted the most important methodological guideline for doing mathematics. First of all, it enabled mathematicians to define concepts such as proof, truth and exactness. The last one of these notions has especially intrigued mathematicians since ancient times. Throughout history mathematicians have repeatedly raised the question: what does exactness mean? In attempting to provide an answer mathematicians have constantly reshaped their science to meet more appropriate and higher standards of exactness. Perhaps most successful of these attempts has been the Greek proposal the intellectual integrity and rigor of which is impressive even today. One of the most pertinent questions relating to the issue of exactness was: what it means for a mathematical entity to be 'known' or 'given'; and moreover: what it means for a problem to be 'solved', and its solution to be 'found'? Classical geometry provided an answer to these questions: geometrical figures were 'known' or 'given' if they could be *constructed* starting from elements that were considered given at the outset; similarly, a problem was considered solved if the required configuration was geometrically *constructed*.

What were the geometrical constructions like and what was their correspondence with manipulation with numbers? The following table summarizes the correspondences be-

[4] Hintikka & Remes.
[5] Hintikka & Remes.

tween numbers and the operations acting on them. [Bos (2001), 123] (N.B. Numbers: natural or rational positive numbers, and irrational positive numbers in as far as they were used at the time in numerical context.)

Operation	Modern notation	Exact	Change of kind or dimension	Analogous operation(s) on geometrical magnitude
Adding two numbers	$+$	Yes	No	Joining
Subtracting a (smaller) number from a number	$-$	Yes	No	Cutting off
Multiplying two numbers	$\times$	Yes	No	Making a rectangle
Dividing two numbers	/ or $\div$	Yes	No	Forming a ratio; applying a rectangle
Extracting the square root of a number	$\sqrt{\cdot}$	No	No	Taking the mean proportional
Solving a quadratic equation with numerical coefficients		No		Plane constructions
Extracting a higher-order root of a number	$\sqrt[k]{\cdot}$	No	No	Taking several $(k-1)$ mean proportionals
Solving cubic and higher-order equations with numerical coefficients		No		Solid or higher-order constructions
Forming the ratio of two numbers	:		Yes	Forming a ratio

Besides purely mathematical considerations, the Greek idea of analysis used in geometry touched other issues as well and undulated into areas of non-technical theorizing. Especially remarkable is the analogy of geometrical analysis to dialectical reasoning noted by Aristotle. In his *Topics* he explains this as follows:

> There are certain hypotheses upon which it is at once difficult to bring, and easy to stand up to, an argument. Such, e.g., are those things which stand first and those which stand last in the order of nature. For the former require definition, while the latter have to be arrived at through many steps if one wishes to secure a continuous proof from first principles, or else all discussion about them wears the air of mere sophistry; for to prove anything is impossible unless one begins with the appropriate principles, and connects inference with inference till the last are reached. Now to define first principles is just what answerers do not care to do, nor do they pay any attention if the questioner makes a definition: and yet until it is clear what it is that is proposed, it is not easy to discuss it. This sort of thing happens particularly in the case of first principles (*arkhē*): for while the other propositions are shown through these, these cannot be shown through anything else: we are obliged to understand every item of that sort by definition. The inferences, too, that lie too close

> to the first principles are hard to treat in argument. [...] The hardest, however, of all definitions to treat in argument are those that employ terms about which, in the first place, it is uncertain whether they are used in one sense or in several, and, further, whether they are used literally or metaphorically by the definer. For because of their obscurity, it is impossible to argue upon such terms; and because of the impossibility of saying whether this obscurity is due to their being used metaphorically, it is impossible to refute them. [...] It often happens that a difficulty is found in discussing or arguing a given position because the definition has not been correctly rendered. [...] In mathematics, too, some things would seem to be not easily proved for want of a definition, e.g. that the straight line parallel to the side, which cuts a plane figure divides similarly (*homoiōs*) both the line and the area. But once the definition is stated, the said property is immediately manifest: for the areas and the lines have the same *antanairesis* and this is the definition of the same ratio. [...] But if the definitions of the principles are not laid down, it is difficult, and may be quite impossible, to apply them. There is a close resemblance between dialectical and geometrical processes. [*Topics*, VIII 3, 158a31–159a2]

These observations about method are impeccable. They constitute a surprisingly modern account of features that are essential to the 'analytic method' that is the hallmark of twentieth-century philosophy: (i) to avoid sophistry you must define your terms unambiguously and connect inference to inference up to your conclusion; (ii) providing examples from specific sciences. It is therefore not a coincidence that the analytic movement in philosophy emerged in the process of evaluating the forms of rationality and inference deriving from Aristotle. Up to Kant the Aristotelean conception of logic was taken as canonical, and only gradually the limitations of syllogistic were beginning to be made apparent. The impressive development of logic from Bolzano on prepared the ground for a thoroughgoing transformation in the style of philosophizing and the goals that philosophy would set for itself.

Analysis in context: mathematics in the 18^{th} and 19^{th} centuries

The birth of analytic philosophy is often associated with the arousal of interest at the end of nineteenth century of many philosophers and scientists in the considerable developments that took place in mathematics, logic and physics. Perhaps the most important determining factor shaping theoretical thought in the first place was the almost simultaneous maturation of algebraic logic and mathematical analysis. Although this development did not immediately lead to philosophical considerations of substantial content, it prepared the stage for accommodating such considerations as a vital part of investigating the foundational questions in mathematics. The story begins to unfold in France in connection with mathematics and philosophy in the late eighteenth century. The names of Condillac and Condorcet figure here prominently, whose semiotic 'logique' marked an important attempt to systematise the ideas pertaining to the *semiotic* character of language in general and discourse in particular. Condillac's central tenet presented in his

La logique was that "the origin and generation both of ideas and of the faculties of the soul are explained according to this method." [Condillac (1780), title of pt. 1] The gist of this method was the "analysis" of our ideas as originating in simple sensory experiences. These ideas were then to be reconstructed in the process of "synthesis" in such a way that the relations between them were clearly revealed. This conception has its unmistakable precursor in Locke's *Essay Concerning Human Understanding*. The doctrines of Condillac do not constitute a system of logic in a sense of defining a set of inference rules and logical operations. Rather, they are to be thought of as a programmatic expression of a novel approach to studying language and its usage. Indeed, Condillac's work might be conceived as a pioneering work in semiotics, as Grattan-Guinness has remarked. [Grattan-Guinness (2000), 15] Condillac is squarely within the tradition of Port-Royal logic and Enlightenment philosophy, laying great emphasis on language. Condillac's aim seems to have been to specify general rules for language-making that would ensure that the proposed language and the signs that constitute the meaning-carrying elements of that language are defined as precisely as possible in order that ideas could be clearly and unambiguously stated or expressed. He maintained that "the art of reasoning is [thereby] reduced." [Condillac (1780), title of pt. 2, ch. 5] Moreover, he thought that the rules of inference were to be defined separately for each particular language. This makes it understandable why he abstained from presenting syllogistic logic (the only known system of deductive inference at the time) in his book. Indeed, the syllogistic form of reasoning represented for him a paradigmatic example of a system of inference rules that was supposed to govern reasoning independently of the given language in which such reasonings were to be expressed. Such an independence was, then, a feature that Condillac thought could not be attained given the complexity and multiplicity of different systems of language.

The clarity for which Condillac and Condorcet were aiming at with their semiotic "logique" was manifested in mathematics, generally considered as a true *apotheosis* of clear science. Among the brilliant mathematicians that had brought their science to a considerable level of sophistication, Lagrange perhaps deserves the pride of place. He had immersed himself in the studies of mathematics in his youth, preferring the algebraic formulations of mathematical theories the most important one of which was mathematical analysis. His investigations in analysis were to a considerable degree catalyzed by the congenial philosophical atmosphere in which he found himself, leading to the formulation of the Principle of Least Action, for example, that reflected Lagrange's metaphysical aspirations as much as his aptitude for mathematical concept formation. The metaphysical motivation derived in part from the philosophy of Leibniz according to whom our actual world is the best of all possible worlds. Therefore, thought Lagrange, elaborating the thought of Leibniz a little further, the laws of nature can be described in terms of extremal principles. The mathematical theory under which the principle found its proper domain of application was calculus of variations. The algebraic formulation of this branch of mathematics was one of the most important mathematical achievements of Lagrange. It is illuminating to give a more exact account of the underlying principle in mathematical terms.

Principle of Least Action Let us consider a dynamical system (a conservative holonomic one) for which the kinetic potential does not involve the time explicitly, so that the integral of energy

$$\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r} - L = h$$

exists. Supposing AB to be part of a trajectory and CD to be part of any adjacent arc, to the successive points of which values of the time are so correlated as to satisfy an equation of the form

$$\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r} - L = h + \triangle h,$$

where $\triangle h$ is a small constant, we have

$$\int_{CD}\left(\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\right)dt - \int_{AB}\left(\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\right)dt$$

$$= \int_{CD}(h+\triangle h)dt - \int_{AB} hdt + \int_{CD} Ldt - \int_{AB} Ldt$$

$$= (h+\triangle h)(t_1 + \triangle t_1 - t_0 - \triangle t_0) - h(t_1 - t_0) + \left[\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\triangle q_r - h\triangle t\right]_A^B$$

$$= \left[\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\triangle q_r + t\triangle h\right]_A^B.$$

If therefore we suppose that C coincides with A and D coincides with B, and that $\triangle h$ is zero, we shall have

$$\int_{CD}\left(\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\right)dt = \int_{AB}\left(\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\right)dt,$$

which shows that the integral $\int\left(\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\right)dt$ has a stationary value for any part of an actual trajectory, as compared with neighbouring paths between the same termini for which the time is correlated to the coordinates is such a way as to satisfy the same equation of energy. This is called the **principle of Least Action**, the integral $\int\left(\Sigma_{r=1}^{n}\dot{q}_r\frac{\partial L}{\partial \dot{q}_r}\right)dt$ being called the **Action**.[6]

The concept of "action" associated with the Principle of Least Action is an abstract one, not amenable to direct measurement. Its significance in modern physics as a fundamental

[6]The Principle of Least Action originated in Maupertuis' attempt (*Mem. de l'Acad., 1744, p. 417*) to obtain for the corpuscular theory of light a theorem analogous to Fermat's 'Principle of Least Time'. Maupertuis' principle was established by Euler (Addit. II, p. 309 of his *Methodus inveniendi lineas curvas, 1744*) for the case of a singular particle under a central force, and by Lagrange (*Miscell. Taurin.* II (1760–1), *Œuvres*, I, p. 365) for much more general problems.

tool of concept formation and theory construction cannot be exaggerated.[7] To have envisaged such an abstract concept on a basis of metaphysical speculation is a remarkable feat of imaginative thinking on the part of Lagrange. Moreover, these bold theoretical developments inevitably led towards a situation in which the dreaded questions of "limits" and "infinitesimals" could not anymore be avoided. Hence, Lagrange must be seen as one of the most important harbingers of modern mathematics, having incorporated substantial novelties into mathematics in terms of concepts, calculation techniques, and theories in the first place, but also as representing such a radical approach to mathematics that instigated developments that necessitated the foundational inquiries that were to form such a significant part of the mathematics of nineteenth century.

Lagrange himself did not foresee all of these developments, of course. The general concepts of "function", "continuous function" and "analytic function" only gradually emerged in the eighteenth century from the writings of Euler and Lagrange, and they seldom defined these concepts with precision, even failing occasionally to distinguish between them. The reaction of Lagrange against the Newtonian treatment of differential calculus is symptomatic of the level of mathematical rigour at the time. In *Théorie des fonctions analytiques* (1797) he goes on to lament that Newton's method of ultimate ratios, "like that of limits ... which is just its algebraic translation [has] the great inconvenience of considering quantities ... when they cease to be quantities." This passage betrays the fact that even the fundamental notion of derivative, the concept that was central to the investigations of Lagrange, remained controversial. The mathematician that most forcefully set a change in motion with regard to these issues was Augustin Louis Cauchy (1789–1857). His work in the foundations of analysis is characterized by an aspiration to give rigorous definitions to the fundamental concepts. The whole edifice of analysis is constructed synthetically from some basic assumptions together with the definitions. Consequently, he takes the notion of *limit* as the point of departure for his enquiries. In *Analyse algébraique* (1821) he defined the notions of limit and continuous function in a form in which they were used until Weierstrass and Heine presented the modern $\epsilon - \delta$ definition around 1880.[8] Grattan-Guinness has summarized the import of

[7] Indeed, the most efficient applications of the principle may be found in Quantum Mechanics. Especially notable is Richard Feynman's brilliant use of the notion in his path integral approach to Quantum Mechanics which essentially utilizes a stationary-action principle. Another context in which the action principle is required is the derivation of laws regulating various physical symmetries. A famous example is Noether's Theorem which states that to every continuous symmetry in a physical system there corresponds a conservation law (and conversely).

[8] Cauchy's definitions were as follows: [*limit*]: When the values successively attributed to the same variable approach indefinitely a fixed value, eventually differing from it by as little as one could wish, that fixed value is called the *limit* of all the others. When the successive absolute values of a variable decrease indefinitely, in such a way as to become less than any given quantity, that variable becomes what is called an *infinitesimal*. Such a variable has zero for its limit. [*continuous function*]: Let $f(x)$ be a function of [the real] variable x and suppose that this function ... has a unique and finite value for each value of x in a given interval. If, to a value of x in this interval, one adds an infinitesimal increment h, the function itself increases by the difference $f(x+h) - f(x)$; this depends both on the new variable h and the value of x. Given this, the function $f(x)$ will be a *continuous* function of the variable x in the interval when, for each value of x in the interval, the magnitude of the difference $f(x+h) - f(x)$ decreases indefinitely with that of h. In other words, the function $f(x)$ will remain continuous relative to x in a given interval if [in this interval] an infinitesimal increment in the variable always produces an infinitesimal increment in the function itself. [Birkhoff (1973),

Cauchy's elaborations in four items: (1) One of the main theorems of analysis was Taylor's,[9] for the related series of which Cauchy defined the forms of the remainder term, explicating thereby the conditions of its convergence. (2) Presenting theorems in terms of sufficient and/or necessary conditions; in general, stating conditions for the validity of the standard procedures of analysis. (3) The status of logic was raised because of just such conditions as mentioned in (2). Cauchy did not, however, adopt any specific theory of logic. (4) The terminological novelty of calling the subject "mathematical analysis". This must be seen as an unfortunate confusion in terms, because the mathematics was mostly developed by synthetic proofs. [Grattan-Guinness (2000), 67]

How are these developments in the mathematical field reflected in the philosophical thought of the eighteenth and nineteenth centuries? There is a tendency to assimilate the mathematical results into a system of concepts that encompasses a wide range of issues ranging from metaphysics to epistemology, from political thought to ethics and more. Two characters especially distinguish themselves in this respect. Kant and Bolzano are two of the most important philosophers whose ideas can in a certain sense be conceived as having prepared the stage for the analytic turn in philosophy. Bolzano in particular holds a special place within the pre-history of the analytic tradition for he anticipated many of the thoughts that Frege propounded in his logical investigations decades later. Husserl went so far as to call this philanthropic Praguese mathematician and philosopher as "einen der Größten Logiker aller Zeiten." [Husserl (1900), §225] Many have shared Husserl's opinion. Indeed, the main achievements of this man are nothing short of miraculous:

> Bernard Bolzano (1781–1848) was a Catholic priest, a professor of the doctrine of Catholic religion at the Philosophical Faculty of the University of Prague, an outstanding mathematician and one of the greatest logicians or even (as some would have it) the greatest logician who lived in the long stretch of time between Leibniz and Frege. As far as logic is concerned, Bolzano anticipated almost exactly 100 years before Tarski and Carnap their semantic definitions of logical truth and logical consequence; and in mathematics he is not only known for his famous Paradoxes of the Infinite, but also for certain results that have become and still are standard in textbooks of mathematics such as the Bolzano-Weierstrass theorem. Bolzano also made important contributions to other fields of knowledge in and outside of philosophy. Due to the versatility of his talents and the various fields to which he made substantial con-

2]

[9] "Theorems of Maclaurin and Taylor: When, in some interval and for values of θ satisfying $0 \leq \theta \leq 1$ ['inférieurs à l'unité'], one of the [two] expressions (1) $x^n f^{(n)}(\theta x)/n!$, (2) $x^n(1-\theta)^{n-1} f^{(n)}(\theta x)/(n-1)!$ tends to zero ['décroit indéfiniment'] as n increases, then, setting $n = \infty$ in equation [...] one finds that (3) $f(x) = f(0) + xf'(0) + x^2 f''(0)/2! + x^3 f'''(0)/3! + \ldots$. Therefore ... the [infinite] series whose general term is the product $x^n f^{(n)}(0)/n!$ is convergent in the given interval, and its sum is $f(x)$. This is Maclaurin's Theorem." [Birkhoff (1973), 6] Maclaurin's Theorem is a special case of Taylor's Theorem which concerns the approximation of a function f with a series analogous to Maclaurin's evaluated in the neighbourhood of a point a. Hence the remainder term in the more general case is of the form $R_n = \frac{f^{(n)}(a+\theta(x-a))}{(n-1)!}(1-\theta)^{n-1}(x-a)^n$.

> tributions he became one of the last great polymaths in the history of ideas. [SEP: "Bernard Bolzano" [Morscher (2007)]][10]

From the point of view of philosophy, Bolzano's contributions are a veritable treasury of fruitful ideas. But it would be erroneous to associate his significance only with some isolated, albeit very interesting, results in different areas of logic, epistemology and metaphysics. Indeed, much more is seen to be at stake. Some researchers, most notably Alberto Coffa, attribute to Bolzano a role in the development of analytic philosophy that stands second to none when observed from the point of view of the central issues that have defined the character of the modern analytic movement, namely, philosophy of language and semantics. As Coffa writes: "Bolzano was the first to recognize that transcendental philosophy and its idealist sequel were a *reductio ad absurdum* of the semantics of modern philosophy. He was also the first to see that the proper prolegomena to any future metaphysics was a study not of transcendental considerations but of what we say and its laws, and that consequently the *prima philosophia* was not metaphysics or ontology but semantics. The development of these ideas in his monumental *Wissenschaftslehre* and in a variety of other writings established Bolzano as the founder of the semantic tradition." [Coffa (1991), 23] This is, indeed, a line of interpretation which most clearly makes explicit Bolzano's significance for modern philosophy. Moreover, in Bolzano's thought the method of analysis and the idea of analyticity come fascinatingly together. In *Wissenschaftslehre* he gives the following explanation:

> [...] daß es Sätze gibt, die ihrer ganzen Art nach wahr oder falsch sind, wenn man gewisse Teile derselben als veränderlich annimmt; daß aber derselbe Satz, dem diese Beschaffenheit zukommt, wenn es nur eben die Vorstellungen $i, j, \ldots$ sind, die man in ihm als veränderlich annimmt, sie nicht auch dann noch behalte, wenn man andere oder mehr Vorstellungen als veränderlich voraussetzt. Insonderheit ist leicht zu begreifen, daß kein Satz so gebildet sein könne, daß ihm die Beschaffenheit, von der wir jetzt sprechen, bliebe, auch wenn wir *alle* Vorstellungen, aus denen er bestehet, als veränderlich ansehen wollten. Denn dürften wir alle in einem Satze befindliche Vorstellungen nach Belieben abändern: so könnten wir ihn in jeden beliebigen andern verwandeln, und folglich gewiß bald einen wahren, bald einen falschen Satz aus ihm machen. Wenn es aber auch nur ein *einzige* Vorstellung in einem Satze gibt, welche sind willkürlich abändern läßt, ohne die Wahr- oder Falschheit desselben zu stören; d. h. wenn alle Sätze, die durch den Austausch dieser Vorstellung mit beliebigen andern zum Vorscheine kommen, entweder insgesamt wahr oder insgesamt falsch sind, vorausgesezt, daß sie nur Gegenständlichkeit haben: so ist schon diese Beschaffenheit des Satzes merkwürdig genug, um ihn von allen, bei denen dies nicht der Fall ist, zu unterscheiden. Ich erlaube mir also, Sätze dieser Art mit einem von *Kant* entlehnten Ausdrucke *analytische*, alle übrigen aber, d. h. bei denen es nicht eine einzige Vorstellung

[10] The acronym "SEP" refers to the internet portal *Stanford Encyclopedia of Philosophy*, this particular reference being to `http://plato.stanford.edu/entries/bolzano/`, retrieved on September 26, 2009.

> gibt, die sich ihrer Wahr- oder Falschheit unbeschadet willkürlich abändern ließe, *synthetische Sätze* zu nennen. [Bolzano (1963), 230–231][11]

These very ideas will figure prominently in the philosophical discussions of the next century. The notion of analyticity that then emerges is not, however, a direct descendant of Bolzano's idea. Rather, some of Bolzano's underlying assumptions are then by-passed as confusing or unintelligible. Only quite recently a deeper insight into the fundamental logical ideas of Bolzano has been gained through the work of such scholars as Jan Berg and Edgar Morscher. Bolzano's impact on the analytic tradition in philosophy is therefore mainly of a programmatic nature, but it has an indirect systematic influence upon it through the work of Frege and Husserl. It is now apt to investigate in more detail what analysis is thought to consist in. I will do this by concentrating on three main senses of analysis that have figured in modern philosophy.

Three senses of analysis

In the first chapter of his *My Philosophical Development* [Russell (1959)], Bertrand Russell presents his readers the following confession, a sort of philosophical *credo*: "Ever since I abandoned the philosophy of Kant and Hegel, I have sought solutions of philosophical problems by means of analysis, and I remain firmly persuaded, in spite of some modern tendencies to the contrary, that only by analysing is progress possible." [*ibid.*, XX] Russell's statement is generally regarded as an expression of the fundamental stricture of so-called analytic philosophy. But in addition to mere sentimental associations, this statement does not tell us much about what analytic philosophy is supposed to consist in. Indeed, during the last few decades there has been a growing concern about what analytic philosophy is, or what it is supposed to be. Nevertheless, a general contention seems to be that its main sources can be unambiguously identified. In such a classification the following names constitute a minimal set: Gottlob Frege (1848–1925), Bertrand Russell (1872–1970), G.E. Moore (1873–1958) and Ludwig Wittgenstein (1889–1951). The work of these men roughly within the period 1880–1920 is considered as the canon of

[11] "[...] that there are propositions which are generically true or false when certain parts of them are allowed to vary; but that the same proposition itself which possesses this quality when the only ideas that are allowed to vary are $i, j, \ldots$, would not retain it when other additional ideas are presupposed to vary. It is particularly easy to comprehend that no proposition can be constituted in such a way that it maintains the quality that we are currently talking about when we regard *all* the ideas of which it consists as variable. For if we could modify all the existing ideas of a proposition as we think proper: we could make such a transformation in each arbitrary case, and consequently turn it into a certain proposition that is sometimes true, sometimes false. When there exists a *unique* idea in such a proposition that allows arbitrary variation without interfering with the truth or falsity itself; i.e. when all the propositions in which an exchange of these ideas with arbitrary ones is made, are brought forth that are either all together true or all together false, presupposing only that they have denotation: then already is such a character of a proposition, strikingly enough, distinguishable from those in which this is not the case. May I be permitted, then, to call the propositions of this species, by an expression derived from Kant, *analytic*, and all the remaining ones, i.e. those in which there is not a single idea that could be arbitrarily varied without detriment to their truth or falsity, *synthetic propositions*." My translation.

analytic philosophy (the time limits set by the occasions of publication of two very influential books, Frege's *Begriffschrift in 1879* and Wittgenstein's *Tractatus* in 1921). Topically the emergence of analytic philosophy relates to the rebellion of Russell and Moore against the tradition of British Idealism, the main ideas of which essentially derived from Hegel's philosophy. But as the ideas of Russell and Moore developed beyond the point of being mere consequences of a reaction against idealist philosophy, the importance of the writings of Frege became more evident. Wittgenstein's *Tractatus* can most naturally be seen as an outgrowth of a reaction, on its part, to Frege's and Russell's ideas with which it critically engages.

Now, central among all these developments, foremost in connection with Russell's and Moore's reaction against idealism is the emphasis placed on the idea of analysis. But, as is clear to anyone who attempts to acquire an understanding of what that idea consisted in, or what the so-called method of analysis was supposed to be, the writings of Moore and Russell provide no straightforward answers. In fact, they were notoriously unclear as to the meaning of "analysis" in general. Nevertheless, a certain feature in their methodology was conspicuous and can be made explicit within a general scheme of classification that Michael Beaney has introduced.[12] In the philosophical work of Moore and Russell, the *decompositional* conception was undoubtedly dominant. What does this characterization signify? Simply put, it means that analysis was somehow conceived as a process of decomposing something into its constituent parts. A representative example of this style of analysis can be found in Moore's early paper "The Nature of Judgement" (1899). In this paper Moore still advocates a naïve realist conception according to which the world is composed of 'concepts' which are then synthesized to propositions. Both of these entities are independent of us. The essence of Moorean analysis is summarized as: "A thing becomes intelligible first when it is analysed into its constituent concepts." [*ibid*, 8] Russell adopted both Moore's commitment to naïve realism and his method of decompositional analysis. But as was soon becoming evident for Russell, the simplistic approach of Moore could not hold water. The resulting philosophical problems lead Russell quickly to switch positions and finally to introduce a position he called logical atomism. Underlying this atomism was a train of ideas deriving from his involvement for some time with the idea of denotation and denoting concepts which notoriously led to some recalcitrant problems. Around the time that Russell presented his views of this "logical atomism", Wittgenstein had already arrived at similar conclusions about the nature of language and its relationship with the world. It was, in fact, Wittgenstein who influenced Russell in his adoption of atomism the ingredients of which Russell picked up in a personal communication with Wittgenstein. Wittgenstein's *Tractatus* advocated a logical atomism that was, however, in many respects different from the one presented by Russell. It might seem like the atomistic philosophy of these two titans of the analytic school was something entirely new. Such is not the case, however. The essentials of the atomistic idea of constitution of things was already present in Leibniz's philosophy. Indeed, his idea of the *characteristica universalis* together with the elaborate metaphysics of *Monadology* can be seen as a form of atomism *par excellence* (although these tendencies were counterbalanced by more organicist and holistic ideas). Furthermore, decompo-

[12] I will present this classification shortly.

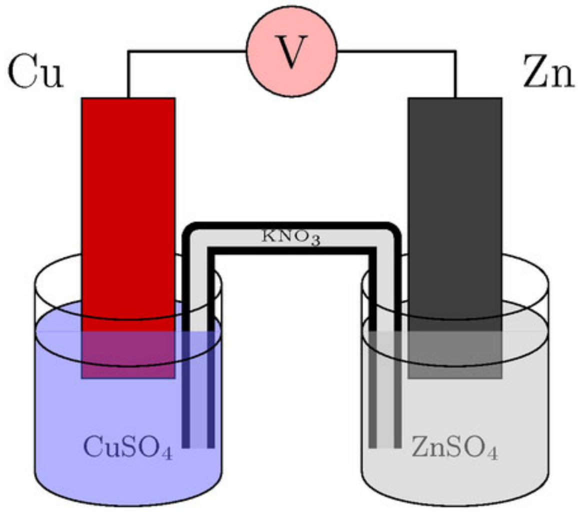


Figure 3.1: Daniell's Pile. ©Agustin E. Bolzan. Source: LaTeX-Community.org.

sitional methodology in philosophy figured prominently in the work of Descartes and Locke, for example, and the classical definition of analyticity provided by Kant remain a hallmark of its application. One idea that gradually became to be associated with this mode of analysis related to analogous procedures in a completely different domain. As a matter of fact, the decompositional variety of analysis bears a conspicuous resemblance to the methods of analytical chemistry. Indeed, the decomposition of concepts to their atomic constituents is *semiotically* very much like the decomposition of chemical elements in a Daniell pile for example, where the chemical compounds copper(II)sulphate and zinc sulphate dissolve in an electrolysis. The chemical reaction is

$$Zn(s) + Cu^{2+}(aq) \rightleftharpoons Zn^{2+}(aq) + Cu(s).$$

This analogy between conceptual analysis and chemical analysis captures succinctly the fundamental presuppositions underlying decompositional analysis. As we have seen, this variety of analysis was prominent already in the early modern period within philosophy. It cannot therefore be a unique characteristic of analytic philosophy. So why is analytic philosophy thought to start with Russell and Moore?

Michael Beaney has answered this question by proposing that it is not the decompositional mode of analysis alone that characterizes analytic philosophy, not even within the severely restricted confines of logical atomism during the early 1920s. Indeed, Beaney suggests that:

> [...] the single most significant event in the development of analytic philosophy was not Russell's and Moore's rebellion against idealism, but the appearance in 1905 of Russell's theory of descriptions. Frank Ramsey rightly described this theory as a "paradigm of philosophy" [...] [Beaney (2007), 2]

As Beaney goes on to explain, what is crucial about the theory of descriptions[13] is that it introduced a completely different conception of analysis which Beaney characterizes as a *transformative* or *explicatory* conception. The fundamental feature of the theory is the *rephrasing* of a given sentence that is to be analysed into a quite different form. To provide a time-honoured example, consider the sentence: "The present king of France is bald". This sentence is analysed as "There is one and only one king of France, and whatever is king of France is bald". What is distinctive about this sort of analysis is the manner in which it seems to 'analyse away' certain concepts or phrases that appear in the *analysandum*. So it forms a marked contrast to the decompositional variety of analysis where the constituents of the *analysandum* are conserved, and indeed, made explicit. Altogether, the transformative variety of analysis was not new either. It has precursors in ideas of medieval logic, in Aristotle's theory of dialectical inference, and ancient Greek geometry. A particularly illuminating description of the essentials of this idea is provided in the articulation by Jeremy Bentham of the conception of paraphrasis. In his *Essay on Logic*, Bentham wrote:

> By the word paraphrasis may be designated that sort of exposition which may be afforded by transmuting it into a proposition, having for its subject some real entity, a proposition which has not for its subject any other than a fictitious entity. [Bentham (1843), 246]

The context of Bentham's discussion was the application of the method in 'analysing away' talk of obligations. [*ibid.*, 247] The similarities between Bentham's and Russell's approaches have been pinpointed by John Wisdom in a book that particularly focuses on this very relationship. [Wisdom (1931)] Now there arises the question: what distinguishes Frege's and Russell's use of transformative analysis from earlier uses? The answer lies in part in the role attributed to quantificational logic. As is well known, this variety of logic emerges only with the works of Frege and Peirce. Quantificational logic provides philosophers with a completely new set of tools to work through their analyses. In the first place, it provided an unforeseen expressive power, enabling one to represent propositions and inferences in a form that was both lucid and systematic. The representations given in quantificational logic provided glimpses to the fine structure of linguistic expressions, propositions and inferential patterns. These enormously facilitated analysis. But this mode of representing propositional contents and inferences resulted also in some severe philosophical difficulties. Foremost among these was the question about the relationship between ordinary language and formal logic. It now became necessary to find criteria that governed the 'correctness' of a logical formalization. At this early period philosophers inspired by these new tools were led to ponder over what was preserved in such transformations. The essential question was: what *is* preserved and what can be

[13]In a nutshell, what is involved in Russell's theory of description is a general scheme for analysing propositions of the form "$G(\iota x)(Fx)$". On Russell's account a proposition of a type which can be expressed by utterances of the form "F is G" is logically equivalent to the proposition expressed by the conjunction of the following three clauses: (a) There is at least one F, (b) There is at most one F and (c) Everything that is F is G. Now, the proposition "$G(\iota x)(Fx)$" is defined as 'meaning the same as' "$(\exists x)((\forall y)(Fy \equiv y = x)\&Gx)$", or equivalently, "$(\exists x)(Fx\&(\forall y)(Fy \supset y = x)\&Gx)$".

allowed to vary? In this novel context such questions could be addressed in a far more detailed form that before. In particular, the coherence of the representations was now dependent on the coherence of the logic itself. Thus, the new approach to philosophical analysis by means of the tools provided by mathematical logic aroused the need to justify the new logic. It is precisely this issue that gives the so-called analytic turn its philosophical momentum. It is arguably true that the analytic turn instigated the linguistic turn that was to become the hallmark of twentieth-century philosophy. Indeed, without the investigations of Frege and Russell into the possibilities of transformative analysis that resulted in the varied applications of quantificational logic, the philosophical questions that have given shape to the analytic tradition at large would largely have remained unasked. The research line of transformative analysis in philosophy ultimately leads to the emergence of a completely new conception of analysis, the program of rational reconstruction, largely associated with the Vienna Circle, and Rudolf Carnap in particular. Rational reconstruction is "the searching out of new definitions for old concepts", where the new definitions "should be superior to the old in clarity and exactness, and above all, should fit into a systematic structure of concepts." [Carnap (1961), v] It is the development of this conception of analysis, that later came to be called "explication", that forms one of the focal topics of this dissertation.

We have so far investigated two different senses of analysis, the *decompositional* and the *transformative* senses. These terms have been introduced by Michael Beaney [Beaney (2007c)] as a part of a classification that comprises *three* different variants of analysis. The third, and final one, is called *regressive* analysis. Why is it not mentioned here in connection with the developments in analytic philosophy? The answer is that it can be identified with the idea of analysis that we have already encountered: the analysis of the ancient Greek geometers. To reiterate, according to Pappus:

> In analysis we suppose that which is sought to be already done, and we inquire from what it results, and again what is the antecedent of the latter, until we on our backward way light upon something already known and being first in order. [Hintikka & Remes (1974), 8]

The conception articulated here has been, and remains today, a core conception of analysis. We have seen examples of it in Aristotle's *Topics*, Descartes' *Regulæ* and Newton's *Principia*. It is inextricably intertwined with the issue of inductive inference, as we have seen in connection with Newton. It has also inspired the notion of abductive inference introduced by C.S. Peirce. Although it is not often explicitly referred to in contemporary philosophical discourse, it nevertheless remains a fundamental presupposition of analytic methodology. Let us turn to the historically important relationship between the notions of *a priori*, analyticity and the method of analysis.

3.2 *A priori*, Analyticity and the Method of Analysis

The question about the nature and origin of the methods of analytic philosophy has been a constant theme in the philosophical discussions during the last decades. A considerable amount of scholarly effort has been put into understanding the fundamental problems, methods and tenets of analytic philosophy. More recently, a historical approach to these questions has acquired momentum, and the history of analytic philosophy is presently acknowledged as an important subdiscipline of philosophy (concentrating essentially on the origins of the analytic movement and its intricate relationship with the continental traditions of phenomenology, hermeneutics and existentialism).[14] Very little, however, has been done to elucidate the relationship between the method of analysis and the concept of analyticity. The latter has played an important role in philosophy since Kant's *Critique of Pure Reason* (1789), although already Leibniz had recourse to it in his distinction between the truths of reason and the truths of fact. The distinction between analytic and synthetic, which Kant brought to prominence, has mapped out the whole sphere of discourse associated with the fundamental metaphysical and epistemological problems in the last two-hundred years. Depending on the context, the distinction has been regarded as "classical" and "vast",[15] or, in a more critical vein, as the most idiosyncratic dogma of modern philosophy (Quine and his interlocutors). In any case, philosophers who have been intrigued by the questions Kant raised, have felt compelled, in one way or another, to define their position with respect to analyticity. The so-called "semantic tradition" from Kant to Carnap, as meritoriously depicted by J. Alberto Coffa [Coffa (1991)], comprises a group of thinkers the works of whom represent a gradual evolution of some of the most important concepts of modern philosophy, including meaning, truth, logical consequence, etc. The notion of analyticity plays a crucial role in this development. Historically the notions of analysis and analytic sentence are intimately related, although this conceptual affinity is not always consciously appreciated among earlier thinkers. A hint of this affinity can already be perceived within the work of such figures as Wolff, Baumgarten and Crusius, and it is even more apparent in the detailed and critical elaborations of Locke and Hume. But it is not until Kant makes the famous distinction that the full implications of the connection start to unfold and become gradually disclosed. The connection between the classical method of analysis and analytic sentences (propositions, judgements, arguments) is still very much alive in Kant's treatment of analyticity. Originally analytic sentences were truths which could be established solely by means of analysis. There is ample textual evidence that Kant was familiar with the traditional method of analysis. This is reflected in the manner in which the terms "analysis" and

[14] A cornucopia of treatises have been devoted to these issues, the most prominent of which are [Dummett (1993)], [Hacker (1996)], [Stroll (2000)], [Baldwin (2001)] and [Soames (2003)]. More detailed presentations of specific aspects of analytic philosophy are [Hylton (1990)], [Stadler (1997)], and [Hanna (2001)]. The most recent monograph combining the enquiry of thematic and historical problems is [Glock (2008)]. Moreover, several collections of essays on the history of analytic philosophy have appeared, for example, [Bell & Cooper (1990)], [Monk & Palmer (1996)], [Glock (1997)], [Tait (1997)], [Biletzki & Matar (1998)] and [Reck (2002)]. Philosophers permeated with Spenglerian spirits might find it appropriate to refer in this connection to Hegel's famous dictum that the owl of Minerva takes flight only at dusk. If so, analytic philosophers are chasing after gaunt specters.

[15] *Prolegomena*, §§3, 5 (pp. 270 and 276 of volume 4 of the Academy Edition of Kant's Works).

"synthesis" appear in his critical writings earlier than the corresponding adjectives, as well as in his feeling the need to explain them in some detail. [Hintikka (1973), 124]

The conceptual background of analyticity

Joëlle Proust starts his lucid treatise *Questions of Form* by quoting Eberhard: "The distinction between synthetic and analytic judgements is not new; it has been known for a long time".[16] [Proust (1989), xi] Proust is drawing our attention to the way in which Eberhard tries to downplay the originality of Kant's distinction in the Critique of Pure Reason by keeping it in the traditional setting of dogmatic metaphysics. But this is a serious mistake. Although the distinction was made by others before Kant — by Leibniz, Wolff and Baumgarten among others — they did not consider it important enough to highlight it. Kant was able to bring about entirely new problems with his distinction. He revised drastically the conceptual background and elucidated the presuppositions on the basis of which new questions became possible. This was his notable accomplishment. But what is the dichotomy actually about? What was novel in Kant's way of formulating the distinction? Let us start by looking at the answer given by Kant himself. In *Prolegomena*, §2 he says:

> [**(A)**] Metaphysical knowledge must contain nothing but judgments a priori; this is required by what is peculiar to its sources. But whatever the origin of judgments and whatever the kind of their logical form, there is a difference between them as to their content, according to which they are either explanatory and add nothing to the content of knowledge, or enlarging in that they increase the given knowledge; the former can be called analytic judgments, the latter synthetic judgments.

In the next paragraph Kant elaborates the distinction further:

> [**(B)**] Analytic judgments say nothing in the predicate that was not already thought in the concept of the subject, though not so clearly and with equal consciousness.

This latter remark has often been taken as the explicit *definition* of analyticity. Especially Quine seems to think in this vein in the "Two Dogmas", where the sole characteristic of analyticity, as traditionally conceived, is claimed to be the one given in (B). But this is a grave misunderstanding at best, and an irresponsible misrepresentation at worst. The essential features of the distinction are given in the passage (A). What needs to be emphasized here is that the definition (A) can be applied to any kind of assertions irrespective of their logical form, whereas the latter quite explicitly hinges on the presuppositions of Aristotelean logic according to which the structure of propositions is ultimately

[16] Proust is quoting Eberhard (Philosopohisches Magazin (Halle) I–III).

analysable in the subject-predicate form. Strictly speaking, the analysis depending on the Aristotelean approach should not even budge from the *inference patterns* built into the logical theory, thereby imposing severe restriction also on the *relations* of analytic propositions (sentences). This feature is crucial if one attempts to elucidate issues such as the analyticity of logical truths, or better yet, define analyticity in terms of inferential relations. Indeed, Aristotelean syllogistics consists effectively of a theory of inference patterns among quantified sentences being generally of the form QXY, where X and Y are universal terms (usually 1-term predicates) and Q is a quantifier belonging to the set $\{$all, some, no, not all$\}$. The syllogisms that Aristotle presented as part of his *Organon* were of the following general types:

$$\frac{\begin{array}{c}Q_1ZY\\ Q_2XZ\end{array}}{Q_3XY}$$

$$\frac{\begin{array}{c}Q_1YZ\\ Q_2XZ\end{array}}{Q_3XY}$$

$$\frac{\begin{array}{c}Q_1ZY\\ Q_2ZX\end{array}}{Q_3XY}$$

$$\frac{\begin{array}{c}Q_1YZ\\ Q_2ZX\end{array}}{Q_3XY}$$

Kant notoriously tried to defend in an essay of 1762, *Von der falschen Spitzfindigkeit der vier syllogistischen Formen*, the thesis that all syllogistic reasoning depends on the single principle *Nota notae est nota rei ipsius*.[17] But the attempt to found logic on a single rule of inference was fraught with difficulties as already Aristotle had observed. In such an attempt it would be necessary to give an account of the procedure of reduction. Such an

[17] "The mark of a mark is a mark of the thing itself." A scholastic principle which C.S. Peirce, for instance, found as representative of what he called 'leading principles'. The basic idea is that on inferential contexts where one presents what Aristotle called an ἐνθύμημα (*enthymēma*), that is, a non-valid syllogism (conferring the conclusion some probability rather than certainty), then one can always make it valid by adding a premise. Peirce's example is:

$$\frac{\text{Enoch was a man.}}{\text{Enoch died.}}$$

Now, by incorporating the principle "*Nota notae est nota rei ipsius*" we can construct the following (valid) syllogism:

account had been elaborated by Aristotle but Kant seems to have remained ignorant of this. In a later compilation of his lecture notes on logic, edited by B.G. Jäsche and published under the name *Logik*, Kant is not able to say much of interest about the syllogistic or its application. Furthermore, he shows little understanding towards efforts to improve upon the legacy of Aristotle. Of his predecessors work in this direction he was conspicuously doubtful. For example, of Lambert's work in particular he says that it contains nothing but useless subtleties. [Kneale & Kneale (1962), 354] Kant's only motivation to discuss formal logic in the *Kritik der Reinen Vernunft* seems to be the attempt to distinguish it from the transcendental logic that formed such a central part of the whole of his philosophical program. In fact, a curious tension is built between Kant's general view on logic that he adduces in the Preface to the second edition of the *KdRV* and his practice of applying logical notions in the body of the text itself. Perhaps this is made clearer if we quote a passage from the mentioned Preface:

> That logic from the earliest times has followed this sure path [of a science] may be seen from the fact that since Aristotle it has not had to retrace a single step, unless we choose to consider as improvements the removal of some unnecessary subtleties or the clearer exposition of its doctrine, both of which refer to the elegance rather than to the solidity of the science. It is remarkable also, that to the present day it has not been able to advance a step and is thus to all appearance complete and perfect. If some modern philosophers have thought to enlarge, by introducing *psychological* chapters on the different faculties of knowledge, . . . , *metaphysical* chapters on the origin of knowledge or the different kinds of certainty according to the differences in the objects, . . . , or *anthropological* chapters on prejudices, their causes and remedies, this could only arise from their ignorance of the peculiar nature of logical science. . . . The limits of logic are quite precisely determined: it is a science concerned solely with the exhaustive exposition and strict proof of the formal rules of all thought. [Kant (1929) [1787], viii]

In the light of the developments in the body of *KdRV*, this is indeed a remarkable statement. For the *modus operandi* of the whole of Kant's *magnum opus* is the application of

---

Nota notae est nota rei ipsius

$$\frac{\text{Mortality is a mark of humanity, which is a mark of Enoch}}{\text{Mortality is a mark of Enoch.}}$$

But this seems to run counter to the distinction between *Merkmalen* and *Eigenschaften* that Frege advocated in his *Foundations of Arithmetic*, §53. Indeed, as seems *prima facie* plausible, the marks of a concept (such as "*humanity*" in our example) are properties of the objects that fall under the first-order concept in question. Therefore Enoch has the marks of "*mortality*" (such as "*of biological origin*", "*dependent on a steady supply of energy*", etc.) as his properties. But, in contradistinction, the marks themselves are not properties of the concept "*mortality*", they are its conceptual constituents. In sum, mortality is a mark of humanity, but not a mark of Enoch. It is a *property* of Enoch. Therefore the scholastic principle "*Nota notae est nota rei ipsius*" is false. It should read instead "*Nota notae NON est nota rei ipsius*". This observation makes many of the Kantian arguments presented in the *Kritik der Reinen Vernunft* dubious (the ontological proof of God's existence being the most obvious example).

the distinction between *general* vs. *transcendental* logic. Simply put, logic as a pure science concerns either the general or special use of the understanding. Giorgio Tonelli elaborates on this: "The *first* use contains all absolute and necessary rules of thinking in general, disregarding the differences between the various objects of thought, and can be termed elementary logic (*Elementarlogik*). The *second* concerns the rules for thinking correctly about *a certain kind* of object, and can be termed the organon of this or that science." [Tonelli (1994), 60] One thing that is worth noting at this stage is that Kant does not seem to conceive the organons of specific sciences as anything more than instruments for verifying and correcting knowledge. The idea of logic as an instrument (or of the organons as instruments) for extending knowledge does not yet appear in the 1781 edition of the *KdRV*. A more productive task is ascribed to the organons only later, in 1790. A further distinction within general logic is between *general pure logic* and *general applied logic*. The former is solely concerned with principles *a priori* and is conceived as the canon of the understanding and of reason. This is, however, to be understood as pertaining to the *formal* aspects of their use only. It is a doctrine *more geometrico demonstrata* and as such completely *a priori* (cf. the purely formal aspects of Aristotelean syllogistic alluded to above). The latter is concerned with the rules for the use of understanding under subjective empirical conditions. These empirical conditions that are imposed on this use are determined by empirical psychology. These rules do not make a distinction between the objects of the understanding, thus failing to fall under either the canon of the understanding (in the former sense) or an organon of any specific science. Kant regards such rules as a compendium of common sense (*gemeiner Verstand*). Kant goes on to enumerate the objects that the class of these rules accommodates. These objects include attention, the origin of error, the state of doubt, of scruple [*Scrupel*] of conviction and so on. [A 52–55; B 77-79] Kant compares the presuppositions of applied logic to the those of practical ethics "which considers these [moral] laws under all the impediments of feelings, inclinations and passions to which men are more or less subjected, and which never can furnish us with a true and demonstrated science, because it, as well as applied logic, requires empirical and psychological principles." [*ibid.*]

There are marks of convergence of Kant's views on logic towards the end of the 1780s. Hence, his overall conception of logic seemed to have acquired a solidity that reflected a clear comprehension of the different parts of logic on the whole. In the Logic lectures of 1789 (*Logik Bauch*) Kant gave the following characterization of the scope of logic:

> Anmerkung
> ¶1. Die Logik ist eine Doctrin, allein deswegen ist sie noch nicht practisch, sie ist vollkommen nach Gestezen des Verstandes und der Vernunft. Eine Erkentnis, die dies Merkmal an sich hat, heißt vollkommen.
> 2. Die Logik ist eine Kritik, sie enthält nur Gründe einer Dejudication, nicht der Construction; sie urtheilt; ob etwas den Gesetzen des Verstandes und der Vernunft gemäß ist.
> 3. – propadevtisch heißt alles, was zur Vorübung gehört. Logik iste keine Propadevtik des gemeinen Verstandes, sondern sie gehört zweyfach zur Philosophie.

> a. als propadevtic
> b. als ein Theil der Philosophie.|
> Die Logik ist der Uebergang der gemeinen Erkentnis zur Philosophie. Philosophische Erkentnis ist Vernunfterkentnis. Man braucht nicht
> α) seine Sinne zum philosophieren
> β) nicht seine Imagination schwärmen zu lassen. Schwärmen heißt: seine Kraft ohne Regeln in Bewegung setzen.
> Das Denken der Philosophie, ist das Bewustseyn der Gründe.[18] [Kant (1998), 17–18]

Summing up the discussion of Kant's conception of logic so far, we can state the common (supposedly valid) opinion that general logic disregards all content (*Inhalt*) of knowledge (propositions), and is concerned solely with the logical *form* of the relationship of the parts of knowledge (propositions) to each other. The kind of logic that does not abstract from *all* content of knowledge and considers also the *origins* of our knowledge of objects is in polar opposition to the general logic outlined above. The underlying motivation for such a logic would be to discern what in our knowledge is transcendental apart from what is empirical, and would thus constitute an essential core of the critique of knowledge, the pure science of the conditions of its possibility. This, of course, is Kant's central agenda in taking up the notion of *transcendental logic*. The central issue under consideration here, namely, analyticity, does not hinge on the intricacies of Kant's transcendental program, however.[19] The major issues here concern the expressive force of *formal logic* and the inferential relations between propositions. Indeed, even within the confines of Kant's *general logic* there arises the possibility of conceiving some logical laws as synthetic, as Jaakko Hintikka has shown. [Hintikka (1973), *passim.*] I do think, however, that in the end the Kantian division of logic would have to be accommodated, *contra* Hintikka, within any enquiry on analyticity that builds on the presuppositions of Hintikka. This pertains especially to the question of logical truths.[20]

Let me now revert to the Kantian distinction between analytic and synthetic statements. It is arguably one of the most important discoveries in the history of philosophy. Yet the distinction would not amount to anything important, had Kant not in his critical work

[18] "Remark – ¶1. Logic is a doctrine, for which reason alone it is yet not practical; it is complete with respect to the rules of the understanding and reason. Knowledge that is marked by these features is called complete. 2. Logic is a [form of] critique, it consists solely of the foundations for justification, not of those for construction; it makes a judgement; whether something is in conformity with the rules of the understanding and reason. 3. – Everything that belongs to preparatory practice is called propadeutic. Logic is not a propadeutic of common sense, but belongs to philosophy in a double sense. a. as a propadeutic b. as a part of philosophy.| Logic is a bridge from common knowledge to philosophy. Philosophical knowledge is knowledge by reason. Man need not α) philosophize over its meaning β) let one's imagination wander. To wander means: to set the force of [imagination] in motion without rules. The thought of philosophy is consciousness of the foundations." My translation.

[19] This is not self-evident. Although many analytic philosophers would discard the Kantian division as irrelevant for the question of analyticity, this contention is not universal. It may be argued that the Kantian division must be retained to make sense of analyticity (the mode of representation of the distinction being, naturally, along with the question about its meaning, a moot point).

[20] The issues relating to the analyticity of logical truths are tackled in [Hintikka (1973), ch.VII].

— especially in the *Critique of Pure Reason* — tried to disclose the background opposition against which the distinction could be appreciated. As a purely formal distinction pertaining to formal logic, the distinction would be useless. This is because in logic objects are considered in the abstract, without reference to the sources of knowledge about them. The distinction becomes essential in the transcendental realm, or in transcendental logic, because this discipline alone can elucidate the role of the different modes of predication, or modes of acquiring concepts. [Proust (1989), 28] It is this transcendental interest in the distinction that plays the prominent part in Kant's motivations to introduce it in the first place. But of what use is the notion of analyticity, then, if it pertains only to the way in which "the connection between subject and predicate is conceived by identity"? Are we forced to admit that the philosophical import of analytic judgements is encapsulated by the trivial character Locke gave to "trifling propositions"? Not at all. To see this it is essential to note how intimately Kant's thought was still in keeping with the traditional modes of analysis, as practiced in the medieval and renaissance philosophy. Kant very explicitly coordinated his notion of analyticity with these traditional modes. I allude here to the conception that he already held in his *Inaugural dissertation*. In it there's a passage in which he says:

> Analysis, taken in the first sense, is a regress from consequence to ground, but, in the second sense, it is a regress from a *whole* to its possible or mediate *parts*, i.e., the parts of its parts. Both terms, synthesis and analysis, we here use in the second sense. [D, I, §1, note.][21]

The distinction between these two modes of analysis is sometimes alluded to by the notions of *qualitative* and *quantitative* modes of analysis. [Proust (1989), 28.] I think this terminology is misleading. In associating the Greek mathematicians' ἀνάλυσις solely with the qualitative aspects of looking for the premisses of a given judgement or assertion in a sense of an 'upstream' procedure in a direction contrary to that of logical or causal implications, one is overlooking the complementary, equally important aspects of the method. In the Greek geometers' ἀνάλυσις the demonstrative or synthetic movement proceeding 'downstream' was at least as important as the nondemonstrative procedure. Moreover, the various constructions employed in their proofs contained qualitative as well as quantitative aspects; for example, a typical proof in Euclid's *Elements* invariably makes use of a definite sort of construction which explicitly draws on a reductive analysis of a geometrical figure into its composite parts; at the same time, however, the proof is of course also logically built on the axiomatic structure of the whole exposition, deriving new theorems from already proven theorems and axioms. So, in contradistinction to the complementarity of the method of analysis among Greek geometers, the mode of analysis that Kant relied on in his philosophy was that of the resolutive type. In general, analysis may be perceived as going from the conditioned back to the condition, *a rationato ad rationem*, but what Kant mainly retained was the conception of analysis as *a toto ad partes*. [*ibid.*] What counts is a progress in distinction. This means a transition from a state of 'confused' knowledge to that of 'distinct' knowledge marking an increase in knowledge, as

[21] Italics added.

Proust puts it, not *materialiter*, but *formaliter*.[22] [*ibid.*, 29] A final remark about Kant's overall conception of analytic judgements: in a sense, for Kant, analytic judgements are reduced to formal identities. But they are not explicit identities in the way that the "trifling propositions" of Locke are explicit identities. These are merely formal tautologies without any informative content. Kant is not adducing his distinction with such propositions in mind. He thinks that analytic propositions are *implicit* identities. They are not *simple* identities. Somehow the resolution of the original concept to its components conveys information about the original, 'global' concept that was not available before the act of resolution. This conception of analyticity has gained some momentum in the twentieth century, most notably in the attempts of Hintikka to approach analyticity with the conceptual tools of information theory,[23] but on the whole it has not been the most popular account of analyticity. For most modern philosophers analytic propositions are still a trifling business.

Kant's notion of analyticity cannot be properly understood without grasping its relation to the notion of *a priori*. Therefore, something must be said of that notion, even if the clarification that can be provided here is bound to be a rough sketch. I will draw on the succinct characterization of *a priori* by C.D. Broad. The notion emerges in association with the famous problem that Kant presents in the beginning of *Prolegomena*, i.e. "How is synthetic knowledge *a priori* possible?" C.D. Broad offers a straightforward solution to the Kantian problem [Broad (1978), 3-8].[24] He maintains that in answering the question "How are synthetic *a priori* judgments possible?" Kant had recourse to a completely different sense of *a priori* than the one endorsed by his most notable precursors, *videlicet*: Leibniz, Locke and Hume. These three philosophers understood "*a priori*" in terms of what Broad calls *absolute a priori*. This is nothing other than the traditional conception according to which one's knowledge of *p* is *a priori* if and only if one can see that *p* is necessary. Furthermore, one can come to recognise that *p* is necessary either *directly* through inspecting its terms and analysing their meaning relations or *indirectly* by showing that *p* is a logical consequence of other propositions each of which one can see by direct inspection to be necessary. However, Kant formulated another sense of *a priori*, which Broad calls the *transcendental a priori*, that was meant to vindicate judgments that are *both* synthetic and *a priori*. Broad describes the meaning of this determinant (transcendentally *a priori* judgment) succinctly:

> It is a judgment which asserts, with regard to all objects of possible human sense-perception, that they must have certain characteristics, because the latter are entailed by certain very general facts about the way in which human minds work. Kant's transcendentally *a priori* judgments are not judgments of *intrinsically* necessary propositions. If Kant is right, they are judgments of propositions which are *necessary consequences* of certain facts about the human mind; but these facts are contingent and so are their consequences. [*ibid.*, 7.]

[22] Proust refers to Kants Logik, §36.

[23] The notions of *surface information* and *depth information* providing, in effect, a criterion for distinguishing between simple and implicit identities referred to above. See [Hintikka (1973)] and [Rantala (1987)].

[24] This solution is controversial, and I believe, still one of the debated points of Kant's philosophy among most Kant scholars quite independently of Broad's arguments. See for example [Guyer (2006), *passim.*].

The distinction between analytic and synthetic judgements has, of course, mapped out the whole sphere of theoretical thought in the post-Kantian era. This is especially so with respect to mathematics. The literature on the foundations of mathematical knowledge (where the distinction has figured prominently) is formidable and much of it derives its main thrust from the seminal studies of both Frege and Russell. Not all philosophers (or mathematicians, for that matter), however, have considered the point of view provided by Kant as even an acceptable point of departure for their technical discussion. Hermann Weyl, for instance, regarded the Kantian distinction a confusion, much to the detriment of the subject of mathematical logic and philosophy of mathematics:

> Kant's distinction between analytic and synthetic judgements (*Critique of Pure Reason*, Introduction) is so obscurely phrased as to render a comparison with the precise concepts of formal validity in mathematical logic almost impossible. [Weyl (1949), 18]

In contradistinction, Weyl very much appreciated the unique and refined analyses of Husserl in the latter's *Logische Untersuchungen* and gave credit to the definition provided by Husserl:

> Analytic laws are unconditionally universal propositions containing no concepts other than formal. As opposed to the analytical laws we have their particular instances, which arise through the introduction of material concepts or of ideas positing individual existence. And as particular cases of laws always yield necessities, so particular cases of analytic laws yield analytic necessities. [Husserl: *Logische Untersuchungen* II, p.254]

Many mathematicians (and philosophers) have felt similarly about the Kantian distinction and have turned to other philosophical sources for clarification of the notion analyticity. Indeed, it has remained a constant theme of interest in modern philosophy. The distinction between analytic and synthetic judgements (similarly, between *a priori* and *a posteriori* knowledge) acquires a very important function in the philosophy of Rudolf Carnap, as will become evident in the chapters to come. Carnap's attempts to theoretically demarcate the notions of analytic and synthetic knowledge becomes one of the leading lights in his journey from 'quasi-analysis' to 'explication'. Quine's highly critical remarks about the analytic-synthetic distinction are seen to lose much of their philosophical weight through the fact that they are very much directed at a position that was supposedly Carnap's but which Carnap, by the time Quine's "Two Dogmas" appeared, in fact had already abandoned two decades before. The philosophical positions of Carnap and Quine in the 1950s were closer to one another than Quine was prepared to admit (then and afterwards). This hints at the underlying pragmatic presuppositions of Carnap's philosophy that gradually worked their way to the surface and became clearly manifest in the method of explication introduced in *The Logical Foundations of Probability* in 1950.

In this section I have provided a concise account of the notion of analysis in its historical context and hinted at the importance of the method of analysis for modern philosophy. Moreover, I have elucidated the historical relationship between the method of analysis and the concept of analyticity. These issues will be investigated in more depth in the second part of the dissertation, when we concentrate on Carnap's philosophical development and the systematic questions that he tried to answer within the confines of the analytic tradition of philosophy. We still have to examine briefly the significance of *a priori* knowledge for ethics. Then, to round out the chapter, we survey briefly the cultural developments that were conducive to the dissociation of the moral from the domain of rational scientific discourse. This phenomenon constitutes the *akme* of post-modern morality.

3.3 *A priori* Knowledge in Ethics

To revert back to the topic of ethics, it is pertinent to take a closer look at the two fundamental questions, viz. the problem of subsumption and the problem of means and ends, which render the knowledge of empirical necessary for a comprehensive ethical theory. We shall mainly concentrate on the issues raised by the latter, and hence the problem of subsumption shall not be given too much weight in the remainder of this chapter. However, it will be a prominent undercurrent in the second part of the dissertation, where Carnap's method of explication and its relevance for the task of social engineering will be examined from a systematic point of view. Although the traditional question of philosophical anthropology, "What is a person?" is not one that belongs to the ambit of Carnap's explicit philosophical interests, it nevertheless – as we shall see in due time – constitutes an essential ingredient of the more comprehensive program of explication, at least as it is conceived by André Carus.

Let us begin with a question concerning the relationship between *a priori* knowledge and ethics which has played a prominent role since Aristotle, viz. "What is the logical relationship between *Is* and *Ought*?"[25] If we acquiesce to Hume's Law, traditionally given the formulation "purely normative propositions do not follow from descriptive propositions",[26] we can answer that surely, there is an evident discontinuity between *Is* and *Ought*, but this does not exclude the possibility of *mixed syllogisms* which allow a derivation of a normative proposition from a descriptive proposition and another normative proposition. For example, providing a schematic form of such a syllogism, from the normative proposition, "x ought to realize value V" and the descriptive (empirical) proposition "Under conditions C, means M is necessary in order to realize V" logically follows the obligation "x should use means M under conditions C". There is an analogy between

[25]On the Aristotelean conception of practical syllogism, see for example *De Anima* 434a16 ff., *De motu animalium* 701a7 ff., *Nicomachean Ethics* 1144a31 ff., 1146b36 ff., and on the concept of means toward an end, *Metaphysics* 1032b6 f. Noteworthy about the Aristotelean syllogism is that the conclusion is not a proposition but an *action*.

[26]As is well known, Hume was the first philosopher to formulate this idea clearly in *A Treatise of Human Nature* III 1,1.

the relationship of the normative and descriptive components in ethics and in substantial descriptive theories of the sciences; whereas the former combines the knowledge from the two distinct sources of the normative and empirical sphere, the latter combines the knowledge of the logical and empirical elements. The necessity of combining the two resources of knowledge is aptly remarked on by Hösle:

> Even if we were to follow Kant in assuming that these two items of knowledge have no common origin, we would not be able to deny that knowledge in the full sense of the word first emerges fro their collaboration. Analogously, every responsible decision presupposes a combination of normative and descriptive knowledge, and no social science that wants to provide a basis for such decisions can balk at this kind of combination – while being aware that the normative element has to be provided by philosophy. [Hösle (2004), 125][27]

Immediately following this passage, Hösle stresses the significance of the concept of a mixed syllogism for even the clarification of the opposition between ethical absolutism and relativism, and rightly so. It is precisely the complementarity of the major premises and conclusions of the syllogism that enable us to distinguish them, and partly explain the mechanism by which they are come by in the first place. According to the universalist conception of ethics, the major premises of an ethical (mixed) syllogism "are valid, like mathematical truths, always and everywhere (naturally, this does not mean that they are acknowledged in every culture, and still less that people everywhere act in accord with them – though neither of these qualifications puts their validity in question)." [*ibid.*] The alluded complementarity resides in the fact that "*conclusions* of a mixed syllogism are neither always nor everywhere valid: their validity presupposes the truth not only of the normative major premise but also that of the descriptive minor premise, and the latter may well depend on the circumstances." [*ibid.*] The distinction between general moral *principles* and concrete *norms* is relevant here; the former are known only within the domain of pure validities, and it is not realistic to assume that many people succeed in rising to the level of their contemplation; concrete norms, on the other hand, "follow from the principles and the descriptive minor premises, are more comprehensible, and so people will stubbornly cling to them even after the situation has changed. This is one of the great dangers of abstract absolutism." [*ibid.*] The danger of moral relativism, however, is even greater, because "for it both principles and norms change, so that people are left with nothing stable to which they can cling." [*ibid.*] Whereas absolutism causes, according to Hösle less damage than relativism, because its mistakes reside solely within the intellectual realm, "relativism misses the dimension of value that is the heart of the moral." [*ibid.*] We have already seen how the ideals of technical, strategic and communicative rationality characterize the sphere of practical ethics; they embody what can be broadly called means-ends rationality. Whereas technical rationality is grounded in the natural

[27] Hösle makes the additional remark that "[t]he fact that the minor premise is descriptive does not imply that it has to be a posteriori. There are also theoretical a priori synthetic propositions, for example, philosophical social science seeks to derive some characteristics of social systems and their developments from their concepts." [*ibid.*]

sciences, providing answers in the realm of subject-object relationships, strategic rationality is grounded in the social sciences, providing answers in the realm of subject-subject relationships. A paradigmatic example of a discipline that mediates between these forms of rationality is medicine, "since it is concerned with a natural object of a very special kind, the human body; and the transformation of medicine into a pure natural science at the expense of folk psychological knowledge does not do justice to this intermediate position." [*ibid.*]

One particularly relevant sphere of *a priori* knowledge that has to be taken into account with respect to ethical theory is the dyad consisting of decision theory and game theory. Although it is impossible to delve into details concerning these disciplines or their significance for ethics, we can make a couple of general remarks about the specific character of the kind of questions that are involved. In the first place, given the general problem of means-ends analysis within a concrete situation, the question about the legitimacy of a given means in a concrete situations is significantly aggravated by the uncertainty associated with the means, i.e. in situations in which the means will entail the preferred consequence with only a certain degree of *probability* or in situations in which the negative consequences of the means are realized with a certain probability. In the latter case one is dealing with the analysis of *risk*. In the second place, decision theory is completely *amoral*, i.e. it is solely concerned with the criteria of rational decision without taking into account elements of practical reason taken in its Kantian sense; a decision is understood 'rational' if it selects among possible alternative actions (including the omission of any action) the one that is subjectively best for the person concerned. Generally, one speaks of "decision under uncertainty" when the person ("agent" in the characteristically value-neutral language of decision theory) framing his decisions does not know (exactly) what will happen and is not in the position to assign a subjective probability to each of the possible outcomes of his action. In decisions involving risk, the assignment of subjective probabilities to each of the possible outcomes is, in contradistinction to the former case, an essential ingredient of the model. To conclude our short remarks about decision theory and game theory, we may mention some qualifications that are necessary in the application of them under the purview of ethics. Vittorio Hösle brings forth four important restrictions: (1) "for ethics, only a *moral* assessment of the various circumstances is relevant"; (2) "the concept of subjective probability is no less problematic than that of subjective preference"; (3) "for decision theory, omission is only one form of action among others, whereas traditional ethics has always recommended that morally risky actions be avoided in case of doubt, even if the risk of not acting is greater than the risk involved in the action, because one is less responsible for the outcome when one does not do something than when one does it"; and (4) "decision theory [...] greatly overestimates human capabilities of calculation". [*ibid.*, 136–146] It is interesting to note that Rudolf Carnap devoted the last two decades of his career to theoretical issues related decision theory, inductive logic and probability. This is a topic which we, however, cannot go deep into in this dissertation.

3.4 The Roots of Ethical Nihilism

3.4.1 The cultural foundations of modernism in Central Europe

To illustrate the continuity and connectedness of some of the most prominent intellectual trends and ideas that defined the cultural *milieu* of the late nineteenth and early twentieth-century Central Europe, especially Germany and Austria, and to see how these trends gradually led to the growth of ethical nihilism, it is essential to notice how this particular historical era was shaped by two contrary influences. These influences were inextricably linked to the traditions of Enlightenment and Romanticism. Their intricate dialectic in the *fin-de-siècle* Germany (and Austria) constitutes the spiritual force field in which modernism took shape. The modernism that we identify with the impressive cultural achievements of the Middle-European nations was very much defined by the fervent activity in the German speaking countries and their major cities, most notably Vienna, the cultural 'hot-house' of the turn of the century. There the vigor of spiritual activity and the richness of its resources found expression in fields as diverse as architecture (Adolf Loos), music (Arnold Schönberg), literature (Robert Musil), philosophy (Ludwig Wittgenstein), logic (Kurt Gödel), mathematics (Hans Hahn, Karl Menger, Richard von Mises), physics (Ludwig Boltzmann, Ernst Mach, Philipp Frank, Hans Thirring), economics (Carl Menger, Ludwig von Mises), to name the most important examples. It was under these unique circumstances that the philosophy of logical positivism was born.

On the background of our theme lurks the question of what one means by culture. According to a succinct definition, culture is "*a historically derived system of explicit and implicit designs for living, which tends to be shared by all or specially designated members of the group.*"[28] Not to put too much emphasis on this point, I take culture to be an independent part of a three-fold social order, along with several social theorists. Jürgen Habermas, for instance, speaks of a society that consists of three subsystems: (i) socio-cultural, (ii) political-administrative, and (iii) economic. The preference for a three-fold division is shared with other social theorists, for example the sociologists Daniel Bell, Johann Arnason and Nicholas Perlas. One can even find this division as a theoretical background assumption in the artistic and social theory of Josef Beyus. What these different approaches have in common is the ascription of an independent role to the cultural sphere. This implies that society is not viewed as a homogeneous entity characterized by a unified structure and organized according to universal principles. Indeed, a heterogeneous picture of society emerges in which the different spheres identified in the three-fold division comprise various complex and interconnected parts, and their interaction is governed by the intricate dynamics characterized by the fact that the different spheres often develop at their own rate, independently of the other spheres, and this is likely to cause friction in the overall development and functioning of the society. As Bell writes:

> Against the holistic view of society, I find it more useful to think of *contemporary* society . . . as three distinct realms, each of which is obedient to a different

[28]See [Kluckhohn & Kelly (1945), 98].

> axial principle. I divide society, analytically, into the *techno-economic* structure, the *polity*, and the *culture*. These are not congruent with one another and have different rhythms of change; they follow different norms which legitimate different, and even contrasting, types of behavior. It is the discordances between these realms which are responsible for the various contradictions within society. [Bell (1976), 10]

This three-fold division is a quite recent invention. The first writer to have systematically developed this idea is Jakob Burkhardt (1818–1897). Burkhardt, professor in the Basel University, spoke emblematically of the three "*Potenzen*", three historically influential powers, that he identified with state, religion and culture. As a mirror of the times and the prevailing habits of thought, this division betrays the fact that economy was not yet considered as an important factor in the development of a society. In contrast, religion was still seen as possessing an indisputable role in shaping the lives of the citizens to justify its separation from the broadly conceived sphere of cultural influences. Of course, this has all changed now, and one of the characteristic features of the cultural development that we call modernism is the conspicuous transition from an ecclesiastical culture to a secular one. What essentially happened in the turn of the century, in the transition from the nineteenth to the twentieth century, was that the notion of culture became far more flexible than what it had been in the times of Burkhardt. To take one notable example, consider the very influential writings, the novels as well as the theoretical cultural essays, of Robert Musil (1880–1942). There is a conspicuous dualism in Musil's conception about culture. On the one hand, he draws on a classical conception of culture conceived as a regulative ideal that governs those practices of men that lead to technological, artistic and social achievements, i.e., a conception that defines culture as a totality. On the other hand, he draws on a clearly modern conception which focuses on the conflict between culture and nature, a theme that appeared already at the times of Burkhardt, but that surfaces really only in the twentieth century, characterizing in a dramatic way the essential tensions in the modernistic society that ultimately deform its structure and undermine its established frameworks for creative activity and spirituality. In the 1920s and 1930s when Musil wrote his most influential works – especially the insuperable documentary of the *fin-de-siécle Zeitgeist*, the novel *Der Mann ohne Eigenschaften* – he was well ahead of his time, seeing culture as a part of society that stands in opposition to the economic and political spheres.[29] Musil was not alone, however, in defending the autonomy of the strictly cultural. Indeed, Alfred Weber (1868–1958), in addition to advocating a variant of

[29] Another characteristic of Musil's original thought – and one that brings it surprisingly close to Carnap's – was a vision of a kind of a 'science of possibilities' or *Möglichkeitswissenschaft* in the *The Man without Qualities*. This vision builds on the distinction between the "sense of reality" and "the sense of possibilities". The latter is attributable to people who are able to "consider what could just as well have been the case, and not to take *what is* more seriously than *what is not*." [Musil (1931), 16] (Italics added.) Furthermore,

> [t]he consequences of such a creative tendency can be remarkable, of course. Unfortunately it often makes what people admire seem wrong, what they prohibit seem permissible, or both seem indifferent. Such possibility-people [*Möglichkeitsmenschen*] are enfolded, it is said, in a gossameer cocoon of mist, imagination, dreaminess, and subjunctives ... When they are praised, these fools are sometimes called idealists, but this obviously covers only the weaker variant of the species that has no grasp of reality to begin with or sulkily avoids it — the variant in which the missing sense of reality, in other words, is actually a shortcoming. But the possi-

the tripartite division of historical and spiritual influences, maintained that culture represents the "seelische" and emotional sides of humanity, "seine eigentliche Wesenssphäre", of which everything else in life is an expression, a form, a materialization, an image, or a symbol. Weber denied the possibility of a linear development of culture. Instead, drawing on a conception of *Kulturbewegung*, he conceived culture as the history of forms of expression and the search for transcendence.

Broadly speaking, the transition from the nineteenth to the twentieth century can be seen as an intense battle of differentiation, and even fragmentation, the cultural sphere attempting to emancipate from the political and economical spheres that undergo a drastic development themselves, finally leading to the overwhelming importance of these two spheres in the modern society. In the first decades of twentieth century there is a bewildering interplay between the three spheres, making the distinction of sharp boundaries between them very difficult, if not impossible. The national socialists, for example, based their activity in part on the strategy of associating the spheres of culture and politics, doing away with their difference. At the other extreme, illusions that politics could be kept out of culture led to the multifaceted aspirations manifested in the *Sportvereine* and sport movements including the establishment of the Olympic Games and the World Cup in football and of the *Wandervogel* movement (in which Carnap participated during his student years in Jena).[30] These movements had a profound influence on contemporary youth culture and attracted also many adults. The different conceptions of culture that formed the basis for the practical activities of these various cultural movements had to defend themselves against competing views of culture. The greatest challenge for the conception of culture as a subsystem of society — effectively underlying many of the modern conceptions of culture in the beginning of the twentieth century — came, perhaps most prominently, from the ideas of Johann Gottfried Herder (1744–1803). His Enlightenment view of culture as an autonomous and all-embracing totality was one of the most important normative conceptions of culture that influenced the discourse in this vein in the German-speaking countries. But it surely was not the only influential one. Indeed, within this period different conceptions of culture abounded. Perhaps in no other period in history has there been as exuberant collective evaluation among the general public about the meaning of culture and its importance for the society at large. Everybody knew something about it, everybody played a part in it, and everybody took part in it.

One of the most important categorical divisions in the evaluation of the significance of culture and its autonomy within the society was the distinction between *Zivilisation* and *Kultur*. In a first approximation, the first might be characterized as something constructed and artificial, whereas the latter refers to qualities describable as inner, deep and authentic.[31] Already Kant had drawn attention to this difference:

---

ble extends beyond the dreams of mentally deficient persons. It extends also to the aspirations of God that have yet to awaken [*die noch erwachten Absichten Gottes*]. [*ibid.*]

[30]I allude to the significance of this movement for Carnap's development in Chapter **4**.

[31]Thomas Mann constantly draws on this contrast, which by 1918 had already become a *cliché*, in his *Betrachtungen über eines Unpolitischen*.

> We are civilized to the point of excess in all kinds of social courtesies and proprieties. But we are still a long way from the point where we could consider ourselves morally mature. For while the idea of morality is indeed present in our culture, an application of this idea which only extends to the semblances of morality, and is in love of honour and outward propriety, amounts merely to civilization. [Kant (1970) [1784], 49]

The idea of culture as something more subtle than civilization, indeed, as something more closely connected with our inner beings was a fundamental tenet of German spiritual life in the late nineteenth century. Culture was understood in Germany as as the true expression of feeling and self, as the individual's inner values, the *Bildung*[32] of its humanity, for which the individual had to struggle in a way that was true to its inner being. According to Humboldt, while requiring education, culture is "etwas zugleich Höheres und mehr Innerliches, nämlich die Sinnesart, die sich aus der Erkenntnis und dem Gefühle des gesamten geistigen und sittlichen Strebens harmonisch auf die Empfindung und der Charakter ergiesst." The most important quality that has to be discerned in this statement is the emphasis put on the significance of the education of the individual; national differences play no part. It is this cosmopolitan quality that figures prominently in the educational and philosophical ideas of modernism that can truly be said to derive from Humboldt and his Enlightenment predecessors.

Alongside with the cultural influences the social circumstances prevailing in the early twentieth century Germany and Austria played a prominent part in the shaping of the modern mind. In a sense, the outward developments helped to steer the intellectual and moral development in a definite direction. As Voltaire had already remarked, the link between the importance and growth of civilization to the establishment and expansion of town and city was not accidental. The extreme urban growth that Germany and Austria witnessed in the nineteenth century was a necessary condition of the emergence of modernism. Some towns doubled their population between 1820 and 1850 or, even like Berlin, trebled it by 1880. Although the trend was general, it was above all the larger towns which increased disproportionately in size; between 1871 and 1910 the population grew about 50% whereas in the larger towns this figure reached 300%. [Vasold (1996), 47] Indeed, the growth in population was strongly correlated with the flourishing of culture, especially in the so-called cities of modernism such as Berlin and Vienna. The fast growth of the city and its multicoloured and bewildering mixture of civilization and culture, of industry, technology, mass rallies, new art forms, and new means of communication were effects that could not easily be understood by the categories and concepts provided by the philosophies of the past. The urban man was born, and in his wake, modern modes of thought, getting their expression in varieties of symbols that signified to many critics the "Zeitalter der Nervösität".

To provide one important example, a prominent force in the overall constitution of the German intellectual and cultural life was the school of German historical economists

[32]The notion of *Bildung* which is essential for the description of the German culture is also a very prominent feature in Carnap's development. I will therefore postpone a more detailed discussion of it until Chapter 4.

which had a powerful influence on German nationalism, economic and social policy, and social thought generally. This influence was coördinated and enhanced through the Verein für Social Politik, a society established to incorporate the theoretical ideas of its members in the concrete political practices of the Prussian government.[33] In regard to influence, the political economists centered around the Verein were far more important in shaping the socio-political mind-set of enlightened individuals than Marx and Engels, whose ideas were still comparably unknown at the time. Their fate was to remain as marginalized outcasts for most of their lives, a state of affairs that was perhaps in no small measure dependent on the fact that their political pamphleteering and explicitly tendential journalism that were known only to a small minority of similarly oriented revolutionaries obfuscated the more solid and serious philosophical ideas on social structure, economics and history, the celebrated advocates of which they would only later become. The situation changed drastically after the Social Democratic Party became a prominent force in the German Society in 1870. As is well known, they derived their inspiration from the writings of Hegel on dialectics of history, of Adam Smith on economics, and of Niebur and Maine on the history of ancient societies. In methodological terms they were opposed to empiricism, holism, and evolutionary theories of historical change.

Compared with continental Europe the intellectual atmosphere in Britain in the first third of the twentieth-century was much more moderate. There was not a discernible tendency to follow certain fashionable currents of thought among the literate and the intellectual atmosphere in general was tolerant of diverse ideas and ideologies. The exuberant discussions and debates about novel ideas in philosophy enlivening the *café salon*s of Paris or the *Cafeteria*s of Vienna were alien to Britain. Indeed, one could not name any singular thinker whose influence on the public at large was especially notable. Even in the academia new ideas were received with calm and only few exceptions exemplified the panache of the intellectual occupations reminiscent of continental Europe. Isaiah Berlin depicts the situation in Oxford in the 1930s:

> I do not quite know what [is meant] by "intellectual currents". I do not think that one can identify currents of this sort — in effect philosophies of life, dominant ideas, crisscrossing with other ideas, as happened in continental and certainly in Russian universities before the Revolution. I do not think that *animateurs des idées* are a typical English phenomenon. [...] There was no philosopher in England who exercised a general influence on the public in the way that Bergson did in France, or Croce in Italy. There were no philosophers

[33]The general purpose and inception of the Verein is succinctly described by Arthur Mitzman: "To prevent the impoverishment of the traditional middle class, to give the workers some legitimate role in the political and social life of the nation, and to keep them from the agonies of applied Manchesterism, in October 1872 an important group of political economists formed the Verein für Social Politik. The group hoped to accomplish its aims by preparing detailed studies of the social problems raised by the new economic era, and by lobbying among economic conservatives and in the government for reform proposals based on its studies. They thus attempted to substitute themselves for social forces that were the outcome of successful revolutions in England and France, and in so doing they continued a long tradition of academic involvement in the affairs of the Prussian state."[Mitzman (1986), 137]

> whose lectures were attended by fashionable ladies.[34] [Berlin & Jahanbegloo (1991), 6-8]

In addition, in contradistinction to the intellectual atmosphere in Germany, evolutionism and holism, rather than materialism and realism, constituted the dominant forces in the thought of major English evolutionary sociologists — most notable among them being perhaps Herbert Spencer, Sir Henry James Sumner Maine, Sir Edward Burnett Tylor and Leonard Trelawny Hobhouse. They embraced the idea that society could be comprehended as analogous to a holistic organism, deriving their influences most perspicuously from the positivism and organicism of Auguste Comte, the drastic development of biological science, the geology of Lyell, and later Darwin's theory of biological evolution. Despite these rather considerable differences in the leading ideas of their cultural lives, both Germany's and Britain's achievements in the sphere of culture would crucially depend on a common set of presuppositions.

3.5 The Dissociation of Morals from the Domain of Rational Discourse: the Case of Tolerance

How did the dissociation of morals from the domain of rational discourse take place? It is clear that the general cultural trends alone were not a sufficient condition for this turn. Indeed, *intellectual* factors played a prominent role in this process. The positivist philosophy of science that emerged in the turn of the twentieth-century, as is well known, articulated the relationship of the scientific and the moral in terms that greatly aggravated the 'crisis' of human reason. We can only make a few comments here regarding moral non-cognitivism and relativism in particular. A paradigmatic example of the influence of moral relativism is the change in the interpretation of the notion of tolerance. Tolerance was one of the constitutive notions of the program of Enlightenment (both in the 'radical' and the 'moderate mainstream' sense), as we explained above in Chapter **2**. We pointed out the place of tolerance (or open-mindedness) within the Enlightenment classification of *intellectual virtues* and within the broader context of the history of epistemic virtues in general. Interestingly enough, there is a hidden connection between the classical Enlightenment virtues and the epistemic virtues that underpin Carnap's work. In the heyday of the Vienna Circle, Carnap advocated *structural objectivity* as *the* intellectual virtue that scientists should prescribe to. This position is characteristic especially of his original and highly idiosyncratic work *Der Logische Aufbau der Welt*. Later, when the shortcomings of the program of rational reconstruction, a sort of report of which the *Aufbau* is, have become fully evident, Carnap turns to a radically different conception of logical analysis in his *Der Logische Syntax der Sprache* which relies on explicit logical pluralism which is manifested in the "Principle of Tolerance". However, the purely *logico-mathematical* interpretation of tolerance propounded by Carnap conceals the highly

[34] Only Thomas Carlyle has had the opportunity to delight in this privilege in Britain, notes Berlin. [*ibid.*, 8]

relevant *ethical* aspects of this notion. As we have seen, the notion of tolerance was originally very much a notion of practical ethics. It might therefore seem a bit surprising to see the notion applied in a completely different context of language engineering. There is, however, a definite connection between these two seemingly very different usages of the notion of tolerance. The *logico-mathematical* conception of tolerance is an outgrowth of a historical process which helped to shape some of our most cherished criteria of epistemic (scientific) adequacy. The development of the different senses of "objectivity" in the eighteenth and nineteenth centuries, for example, largely went hand in hand with the development of scientific practices and experimental techniques which, in the end, called for a broad-minded view on what particular methods could be regarded as 'scientifically acceptable'. Although the criteria of objectivity were transformed in the process, the proposals that had gone before were never fully discarded. On the contrary, they retained their well-defined niches in the *Pantheon* of epistemic practices. These practices stayed alive because of the various needs of artisans, artists, scientists and other professionals that formed in parallel with the emergence of new scientific instruments and experimental techniques, such as photography. The multifarious practices of these different professions were very much characterized by different "epistemic ways of life". These, in turn, were constituted by various techniques of managing, not just the instruments in the laboratory or in the studio, but also the conduct of the artisan-artist-scientist. Hence the question about objectivity turns out to have as much to do with the fine details of epistemology and methodology as with character formation and ethics. The "history of objectivity" as narrated by Lorraine Daston and Peter Galison [Daston & Galison (2007)] is important precisely because it brings to light the intricate relationship between ethical and epistemological dimensions of objectivity. However, Carnap's philosophical views cannot easily be accommodated within the framework of epistemic virtues taken in the broad *ethico-epistemological* sense. His philosophical style already exemplifies the kind of fragmentation that, for example Husserl warned against in his *Die Krisis der europäischen Wissenschaften und die transzendentale Phänomenologie*, and that was conducive to the kind of deterioration of ethics that became the hallmark of Anglo-American philosophy in the twentieth-century. By way of an example, let us see how the notion of tolerance has been burdened by a semantic inflation within the confines of modern discourse ethics.

3.5.1 The modern dilemma of tolerance: from factual pluralism to an essentially incoherent relativism

Tolerance is a notoriously ambiguous notion. The contemporary philosophical and ethical discourse in which this notion appears is muddled. It seems that "tolerance" does not have a definite meaning. This is betrayed by the lack of any general rules or standards to guide its use within the various contexts in which it is applied. Indeed, the very locutions "X is tolerant", "to tolerate Y", "promoting tolerant policy towards Z" etc., have become extremely vague because of their uses as ideological slogans. In folk-psychological parlance toleration is often associated with an all-around indifference to, or a happy-go-lucky allowance for actions and intentions that disagree with the (local) socially valid mores. However, such a view on tolerance can at worst be tantamount to a thoughtless democratic blunder: allowing everything entails allowing nothing.[35] Obviously, tolerance that allows everything amounts to anarchy, and then, as we have seen, we can only speak about morality, which designates the distance from the collective identity of one's own culture. The morality of 'abstract tolerance', as such a naïve view of tolerance may be called, is dangerous and harmful because it has the consequence of isolating the individual from the concrete values of his mores, and worse yet, from the collective identity that springs from mores of the society. Abstract tolerance is an opprobrium of the notion which, for example, Spinoza propounded and the clarification of which was in the interest of the best minds of the Enlightenment. At worst, the attitude associated with such a harum-scarum notion could even amount to a reversal of traditional morality (a state of affairs that Spinoza very clearly abjured and warned about): even if it did not explicitly urge people to support evil (a not altogether unrealistic possibility), it might accept it and consent to it. Of course, this has everything to do with the problem of the relativity of ethical values, an issue which we have become accustomed to; ethical nihilism as a sort of a state of equilibrium, a *status quo*, within modern western societies.[36] The good has become only one of the options on the scale of values accommodating several equivalents. The issue of choosing between good and evil has thereby been trivialized. (Whether this is only an aspect of man's gradual self-liberation, I leave for the reader to decide. The detrimental effect of the trivialized rhetoric of 'tolerance' has nevertheless, I think, received far too little attention from the side of philosophers.) This shows, first of all, that the notion of tolerance defined in such a vague way cannot be a criterion for *constructing* a system of morality; on this interpretation, it is simply inadequate for the background framework of human action. Indeed, the way that the notion is generally used in ethics is more accurately defined as being a 'forcing principle'[37] for situations where different normative frameworks are in conflict ("every opinion is equally respectable"). For example, in ethics the 'principle of tolerance' enters in a quite vague way as part of arguments having the following structure:

(i) the recognition of the existence of rival *factions*, i.e. actual people who do not agree with each other.

(ii) the recognition of the existence of rival *views*, i.e. abstract positions which are incompatible with each other.

(iii) the positing of a right to respect owed to the *holders* of the various disputed positions on a personal basis, i.e. the political and social principle of toleration.

[35]Or, as Norbert Wiener put it with respect to the liberty of a mathematician constructing new concepts: "To be free to do anything whatever is to be free to do nothing." [Wiener (1954), 62]

[36]The positive effects of such nihilism, like the general tendency to react towards the dogmatic strictures of traditional organized religions, do not compensate for the moral losses that are suffered by modern men and societies because of the confusion of the normative and descriptive levels. The liberation from inadequate mores which is attainable without compromising the autonomy of the moral, is illustrated by the hilarious anecdote told by T.W. Körner, the Cambridge mathematician: "[...] In Cambridge during the early 19th century attendance at lectures was not compulsory but attendance at Chapel was. 'The Choice', thundered supporters of compulsory chapel, 'is between compulsory religion and no religion at all.' 'The difference', replied the opponent, '...is too subtle for my grasp.' " [T.W. Körner, "In Praise of Lectures", 2n]

[37]Cf. the idea used in mathematics, where, intuitively, forcing consists of expanding the set theoretical universe V to a larger universe V^*.

(iv) the positing of an equal right to respect for the conflicting *views* that they hold, i.e. a position often mistakenly seen as a corollary of (iii). [Almond (1998) [Morscher et al. (1998), 279–280]]

In this way many arguments are presented in a form that constitutes a progression via the 'principle of toleration' from factual pluralism to an essentially incoherent relativism. Another type of pluralism emerges in situations where there is no reference to cultural differences or alternative value-systems. This is a type of pluralism often encountered (and implicitly assumed) in discussing moral dilemmas.[38]

What, then, are the prospects of defining tolerance as a respectable epistemic virtue? One possible line of explicating the notion and rendering it respectable might be the following. In as much as tolerance is essentially a *propositional attitude* bearing *ex hypothesi* definite logical relations to such deontic notions as *permissible*, *impermissible*, *obligatory*, *gratuitous*, *optional*, *ought*, etc., it might be possible to explicate it within *deontic logic*. Such

[38]Moral dilemmas may be said to arise either when an individual confronts two conflicting obligations, or when two people have incompatible obligations. First, in the individual case it may be argued that there are situations in which two incompatible courses of action are each morally binding on one person. In the second case, two different individuals may each have a moral obligation in a situation in which it is impossible that both action could be performed. How does this view of moral dilemmas comply with the traditional Kantian view that "a conflict of duties and obligations is inconceivable"? Although some moral philosophers have recently taken the course of arguing that moral dilemmas are possible and that there is room for a reasonable pluralism that shows the compatibility of moral conflicts and objectivity in ethics, (For example, Judith Wagner deCew in "Moral Conflicts and Ethical Relativism", *Ethics*, **101**, 28–41.) I think that this line of thought is mistaken. Indeed, as Almond notes, "for while people do indeed sometimes have difficult choices to make, difficult choices are not genuine moral dilemmas. That is to say, choosing between options which are *not* morally equal is not a *dilemma*; on the other hand, though it may be emotionally traumatising, choosing between moral obligations that are indisputably of equal weight is not a *moral* problem, any more than choosing between two paths of equal length to a desired destination, is a mathematical problem." [*ibid.*] We are reminded here of Buridan's ass: if it had been capable of grasping this latter point, it might not have starved to death between two absolutely equal bales of hay. Hence, we are prevented from pleading on situations usually described as moral dilemmas as justifying a form of ethical pluralism. If the alleged pluralism entails the necessity of rational differences on matters of morality, what are the possibilities for rational consensus in the absence of moral dilemmas of the type described above?
Almond elaborates this further: "As a philosophical position the tenet of the possibility of rational consensus arises simply from the belief that people as independent individual thinkers and agents *could* reach shared conclusions on matters of morals. This entails the possibility for arguing for different positions on grounds that take into account the fact that some positions are better-grounded than others, some conclusions more suspect than others. This grounding must ultimately lie in a fuller understanding of human nature. This understanding itself must lie in the deeper appreciation of the links between facts and values — essentially based on the knowledge of human biology, psychology, history, of political and social organisation, which at least provide preliminary hints about where the underlying universal human interests might lie, and what restriction and limitations of action human beings are prepared to accept in pursuing these interests. This conception of common humanity as a fruitful starting-point for moral and political reflection is not at all new, of course. It springs from the classical Greek conception of man, comprising essentially of two basic notions: (i) the perception of moral values as transcendent – as by-passing the particularities of social and political arrangements, of temporal and geographic locations, and (ii) the principle of the primacy and priority of the ethical in human affairs. The upshot of this view is that the postulation of an internal ethic for a group is simply another form of relativism which stands in the way of a self-confident and positive morality." [*ibid.*] It is just because of failing to realize this dimension of *universality* of the notion of tolerance that the contemporary analyses of its philosophical import have went astray. It is the task for modern Enlightenment to rectify this crooked timber of philosophical thinking.

an approach seems *prima facie* plausible, because intensional logics provide a wide range of possibilities for devising the formal semantics of modal notions. But there are reasons for being skeptical about this possibility. Tolerance is primarily a meta-level concept regulating the conduct of scientific (ethical) discourse on a very general level. It therefore seems doubtful that the propositional attitude approach to tolerance would achieve very much in the way of clarifying the notion of tolerance. Indeed, the importance of the Principle of Tolerance consists in providing the framework in which various proposals for a solution of a specific problem, moral or scientific, can be rationally evaluated. We shall see that Carnap, because of his moral non-cognitivism, articulates merely the scientific variety of tolerance, and even then, because of the intimate link between the moral and scientific, fails to convince his colleagues about the significance of the Principle. This is an example *par excellence* of the negative consequences of the lack of ethical virtue. As we will see, in spite of the notable tension between the moral and the scientific, Rudolf Carnap's contributions in scientific philosophy are second to none; his most important work no doubt belongs to the specific areas of philosophical semantics and inductive logic, but also the architectonic of his thought and the so-called ideal of explication have significant interest for the modern philosopher. Let us turn to examine the development of his thought in detail.

Part II

From Rational Reconstruction to Explication

Chapter 4

CARNAP'S EARLY CONCEPTION OF ANALYSIS: *rational reconstruction*

4.1 Carnap's Education: the Jena Years

The years 1890–1914 comprise arguably one of the most significant periods in Germany's modern history measured by its cultural, social, economic and political consequences. During this short interval the German state undergoes a radical transformation from a loosely connected coalition of 16 *Länder*[1] constituting the monarchic regime led by the *Kaiser* to the centrally governed parliamentary monarchy ultimately destined to erode in the turmoil of world politics. It would be of great interest to chart this development, in outline at least, because Carnap's formative years coincide with this tumultuous era of political and social change. In this dissertation I have to settle, however, with a far more modest account of these phenomena. I will bring forth issues pertaining to the historical and political background, in so far as they are relevant for this study, in connection with the systematic questions whenever they illuminate the problems discussed. In short, I entertain a principle of economy and refrain from providing too many historical details.

4.1.1 Ideological influences

One of the most original features of German culture in the eighteenth and nineteenth centuries, to some extent still visible during *Das wilhelminische Vorspiel*[2] 1890–1914 coinciding

[1] These included, at the turn of the century, the following: Prussia, Bavaria, Saxony, Württemberg, Baden, Thuringia, Hesse, Hamburg, Mecklenburg, Oldenburg, Brunswick, Bremen, Anhalt, Lippe, Lübeck and Schaumburg-Lippe.

[2] This locution derives from [Krockow (1990)]. Christian Graf von Krockow has divided the 100 years of German history before Unification in 1990 to three periods, each which its own, distinctive character as well as definite cultural and political significance. [Krockow (1990), *passim.*] The period from 1890 to 1914 is *Das wilhelminische Vorspiel*, the period of gradual abrasion of the monarchic regime and erosion of the inherited,

with Carnap's formative years, was a form of intellectual and emotional cultivation referred to as *Bildung*. It is very hard to translate this term into English but it conveys ideas relating to "education", "educatedness", "character formation", "shaping", "personal development", and "self-cultivation". Growing organically out of the classical conception of cultured refinement, or, 'civilization', it suggests mainly an idea of mental or spiritual development. Its most representative advocates in the classical period, with Kant and Goethe as the fixed stars of the cultural firmament, were Wilhelm von Humboldt and Friedrich Sleiermacher. Following this tradition there emerged such widely differing thinkers and writers as Arthur Scopenhauer, Adalbert Stifter, Friedrich Theodor Vischer, Friedrich Nietzsche, and Theodor Fontane. In the 'modern' era its central ideas have been propounded in a specifically personalized form by Thomas Mann whose renowned novel *Der Zauberberg* can be seen as an expression of a modern version of *Bildung*.[3] To describe the characteristic modes of thought that accompanied the ideal of *Bildung*, the following excerpt from Humboldt's letter to Schiller illuminates some of the essential ideas advocated:

Everyone must seek out his own individuality and purify it, ridding it of the

classical values of *Bildung* and *Tugend*, preparing the stage for *Das deutsche Drama*, the era of the two World Wars, 1914–1945. The story of the post-war Germany is denoted by Krockow with the unimaginative title *Die Deutschen seit 1945*.

[3]The question of *Bildung* is not altogether without interest even within the confines of most recent pedagogical discourse in Germany. In his thought-provoking book *Bildung im Umbruch*, Jürgen-Eckardt Pleines has enquired the presuppositions and conditions of possibility of devising educational principles suitable for our modern age along the lines of the classical *Bildungsideal*. At the end of his book he formulates three central problem areas or questions that would have to be addressed to proceed further: "1. Man sollte sich gegenwärtig ernstlich fragen, ob man die Bildung oder die Ausbildung von Menschen vorrangig im Auge habe. Denn das macht bei der Beurteilung der eigenen Situation einen nicht unerheblichen Unterschied. Was sich gelegentlich als sogenannte Fortbildungsveranstaltung geltend macht, bleibt doch oftmals weit hinter dem zurück, was sie erreichen sollte, nämlich Aufklärung über diejenigen Bedingungen und Zwecke des Handelns, die wir aus guten Gründen für sinnvoll und für vernünftig halten. Die vorrangige Sorge muß daher darin bestehen, was sich aus einsichtigen und öffentlich vertretbaren Gründen der Zeit gemäß gemeinsam wissen, wollen und schätzen läßt. 2. Die Frage, ob Harmonie für uns heute noch ein sinnvolles und notwendiges Bildungsziel sein könne, hängt wesentlich davon ab, was man unter Harmonie versteht. Dabei sind gemeinhin Erwartungen im Spiel, die man aus geistesgeschichtlichen und aus antropologischen Gründen erst einmal auf den Prüfstand der Vernunft bringen sollte. Jedenfalls sollten wir uns davor hüten, an den 'Träumen eines Geistersehers' fortzuschreiben, von denen Kant gelegentlich warnte. Ob es da im Gegenzug Sinn macht, von 'Konfliktfähigkeit' als dem eigentlichen Ziel aller Bildung zu sprechen, bezweifle ich. Dieses Bildungsziel legt doch die ebenso irreführende Vermutung nahe, Konflikte gäbe es nun einmal und sie müßten als Störfälle oder als Fremdkörper beseitigt werden. Es gibt doch gelegentlich Probleme, die werden nicht gelöst, sondern erkannt und als grund-legend anerkannt. Dazu gehört allemal die *conditio humana*, auf die sich vernünftiges Wissen und Handeln durchgehend verpflichtet weiß. 3. Wenn der Begriff der Harmonie die innere Spannung von Gefügtem und damit zugleich die gegenwendige Bewegung alles Seienden bedeutet, dann läuft die uns gewihnte Rede von einer 'harmonischen Bildung' letzlich auf eine Selbsttäuschung hinaus. Vernünftiges Reden, Denken und Handeln ist der Austrag einer gegenwendigen und um sich wissenden Bewegung, die wir als 'menschliches Leben' jederzeit wahrnehmen und geltend machen. Deshalb sollte man in der Vorstellung wie im Denken sehr scharf Harmonie und Symphonie unterscheiden. Wer aber sein Leben zu einer Symphonie, zu einem wohltuenden Zusammenklang gleichartiger Töne gestalten will, wird scheitern. Denn Leben ist in all seinen Formen und Phasen *harmos*, Wiederstreit dessen, was sich in der Spannung fügt. Auf eben diese Grundsituation eines um sich wissenden Lebens, das sich in Kunst und Wissenschaft manifestiert und wiedererkennt, will Bildung aufmerksam machen. Und von einer solchen Bildung sollte hier Rede sein." [Pleines (2000), 259–260]

> fortuitous features. It will still be individuality, for a portion of the fortuitous is inseparable from the make-up of every individual, and cannot an should not be removed. It is really only in that way that character is possible, and through character, greatness.[4]

The core of Humboldt's conception of individual essence was modeled on the examples of Goethe's Wilhelm Meister and Goethe himself. He did not want his character to be a product of chance, although he was fully aware at the same time that he could never be, nor did he wish to be, other than himself. In both a symbolic and romantic sense "Bildung" meant for him the weeding of his mental and emotional garden, resembling the Ciceronian *cultura animi* or the classical hellenistic conception of *paideia*.[5] Henri-Irènèe Marrou has given expression to this same idea in his *Histoire de l'education dans l'antiquitè*, paraphrasing the idea of Plotinus's phrase "modeler sa propre statue" in the following terms:

> Se fair soi-même; dègager de l'enfant qu'on a d'abord ètè, de l'être mal dègrossi qu'on risque de demeurer, l'homme pleinement homme dont on entrevoit la figure idèale, telle est l'oeuvre de toute la vie, l'oeuvre unique á laquelle cette vie puisee être noblement consacrée. [Marrou (1948)]

This passage is an accurate appraisement of the attitude that Humboldt propounded. Indeed, Humboldt summed his own views pertaining to conduct and ethics in the following 'maxims': "The first rule of a true ethical code is 'Improve yourself', and 'Influence others through what you are' comes only second." This kind of an individualism does not form an easily assimilable attitude to life. This was especially the case in the late eighteenth century (and even more so today), but for Humboldt it was a result of a firm ethical decision that was to be reinforced later through the influence of his eminent friends like Schiller, Goethe and Körner. In another letter to Schiller in 1796, Humboldt laid out the ultimate desiderata of a life built on the notion of *Bildung* that he had envisaged:

> If we imagine a man whose sole aim in life is to cultivate himself, his intellectual activity must finally be concentrated on discovering (a) *a priori*, the ideal of humanity, and (b) *a posteriori*, a clear picture of mankind in reality. When both are as precise and complete as possible in his mind, he should, by comparing them, derive from them rules and maxims for action. [*ibid.*, 277f]

[4] *Briefwechseln zwischen Schiller und W. v. Humboldt* [1900], 3rd edition, with notes by A. Leitzmann (Stuttgart), 176.

[5] Werner Jaeger has elaborated on the manifold meanings and pragmatic reverberations of this notion in the classical Greek culture. Jaeger argues that through Socrates this notion has come to encompass problems that are at the very root of our modern ethical outlook, giving emphasis to the notion of "self-mastery": "Der Begriff 'Selbstbeherrschung' ist durch die Sokratik ein Zentralgedanke unserer ethischen Kultur geworden. Er faßt das sittliche Handeln bereits als etwas im Innern des Individuums Entspringendes, nicht nur als die äußere Unterwerfung unter das Gestetz, wie es der herrschende Begriff der Gerechtigkeit forderte. Aber da das ethische Denken der Griechen von dem Gemeinschaftsleben und von dem politischen Begriff der Herrschaft ausgeht, so erfaßt es den inneren Vorgang ducrh die Übertragung des Bildes einer wohlregierten Polis auf die Seele de Menschen." [Jaeger (1954), 103]

As much as the overall significance of the classical ideal of *Bildung* was being mitigated towards the twentieth century, its central tenets still markedly influenced the generation of Carnap, and perhaps Carnap himself more than anyone else among his peers, because the educational ideas behind the *Bildungsvollendung* were entertained by his maternal grandfather, the eminent educational thinker, Friedrich Wilhelm Dörpfeld (1824–1893). Dörpfeld's daughter Anna, Carnap's mother, wrote a memoir of his life, and Carnap's first reminiscences of literary work relate to the book written by his mother. Carnap summed up his feelings in 1963 by relating that he was "fascinated by the magical activity of of putting thought on paper", seeing in that experience a premonition of his life career, adding that "I have loved it ever since." [Carnap (1963), 3]

Carnap and the *Bildungsideal*

It is essential for the purposes of presenting the main thesis of this work to delve in some depth to the background of this German cultural context. I will argue that the overall philosophical worldview of Carnap and his entire philosophical program culminating in the ideal of explication can be best interpreted as an outgrowth of the particular impulses stemming from the *Bildungsvollendung* or *Bildungsideal*. As will become clear later, Carnap considered the thoughts underlying the programme of Vienna Circle as forming a basis for the attempt to constitute a new comprehensive ideal of civilization. This ideal was to be presented as an alternative to the inherited German forms of social organization that had caused such devastating catastrophes as the total political and social failure of the Weimar Republic followed by the First World War. This ideal he shared with Otto Neurath, a zealous advocate of radical ideas of social renewal with leftist political leanings. While Carnap at some point, especially during the World War, promoted the idea of political involvement in tackling the social problems of the German society, and even participated in radical politics, he ultimately came to think that such activity remained inevitably at the surface. To bring about a radical change in society and its structure in a democratic post-war society, a more robust intellectual framework was needed as a basis. This basis was effectively, a new *Bildungsideal* adequate for the modern era. It is this search for the fundamental ingredients of such an ideal that forms a significant part of Carnap's early intellectual strivings. Rather than remaining merely at the level of an expression of a personally experienced *Lebensgefühl*, it is an essential component of Carnap's philosophical temperament, or as I will argue, an indispensable element in his construction of his mature philosophical program, the ideal of explication. I will now delve in more depth to the exemplars of Carnap's early thought where the *Bildungsideal* figures prominently.

Friedrich Wilhelm Dörpfeld (1824–1893) was heavily influenced by the philosophical tradition stemming from Friedrich Herbart (1776–1841). Herbart's philosophy was in part characterized by opposition to German idealism in general, and the positions of Fichte and Hegel in particular which, according to him, left no room for social influences. Ironically, Herbart was originally a member of a group of thinkers which advocated absolute idealism. In the fervent years of idealism, there was a total of three such groups, each with its own peculiar approach to idealism, but all of them offshoots of early romanticism or what could be called *Frühromantik*. The first of these groups was the "romantic circle" in Jena and Berlin consisting of Friedrich Schlegel (1772–1829), Friedrich Wilhelm Joseph Schelling (1775–1854), and Friedrich von Hardenburg (1772–1801).[6] The second group was called as *Bund der Geister* which met in Frankfurt am Homburg, consisting of a circle of friends comprising Friedrich Hölderlin (1774–1843), Georg Wilhelm Friedrich Hegel (1770–1831), Isaak von Sinclair (1775–1815), and Jakob Zwilling (1776–1809). The third group was called *Bund der freien Männer*, meeting in Jena and modeled on Fichtean principles. Its members included August Ludwig Hülsen (1765–1810), Johann Erich von Berger (1772–1833), Johann Smidt (1773–1857), Johann Georg Rist (1775–1847), Johann Casimir Böhlendorff (1776–1825), and Johann Friedrich Herbart. In the autumn of 1794 the philosophy of Fichte was coming under heavy criticism in Jena. This criticism was propounded by those pupils of Reinholdt who had come to question the presuppositions of foundationalism. Herbart was among those critics. He saw Fichte's philosophy as a desperate attempt to rehabilitate an approach to philosophy that was severely undermined by the alleged impossibility of creating complete philosophical systems after the manner of the *Wissenschaftslehre*. The gist of Herbart's novel approach, in addition to the negative import of criticizing Fichte, was his insistence on the social situatedness of man. The social circumstances surrounding man and the interactions between him and his social environment were crucial in character formation. Especially the influences assimilated in childhood and early youth were regarded by him as decisive. In another vein, he stressed that philosophy in general and metaphysics in particular could be put on a solid foundation solely by giving up the idealistic exaltation characteristic of Kant. Indeed, the results and knowledge acquired by the natural sciences should be taken as the starting point of every philosophical enquiry. As a vital part of his overall program Herbart envisaged a mathematical psychology that could rigorously analyse the various factors affecting human action and deliberation. In such a psychology the conscious events of human mind were seen as epiphenomena over the true constituents of mental activity, the 'reals', envisioned as analogous to the Leibnizian monads, exerting various effects (forms of 'pressure') on each other. In accordance with the ideals of *Bildung*, Herbart thought that individual self-development implemented through active interest in 'higher' knowledge and art was an absolute good in itself. Thus, as Carus has maintained, Herbart could well be seen as a representative of the more traditional form of the Enlightenment values in Germany at a time when the intellectual *milieu* was far more receptive and conducive to the romantic ideals advocated by the absolute idealists. [Carus (2007), 43]

Dörpfeld remained attracted to the ideas of Herbart, and in a certain sense promoted those ideas in his own pedagogical writings and teaching. Moreover, in the field of educational politics he resisted the bureaucratisation and centralisation of Prussian primary and secondary education. This was a distinctively Herbartian tenet. As regards his pedagogical thinking, he was always aiming at explaining the task of a teacher in a scientific manner, enabling a practising teacher to understand his duty according to reasoned principles, always asking *what*, *how*, and *why*. Therefore he unceasingly emphasized the supportive role of philosophy, logic and ethics in the work of teachers. He even con-

[6] Better known by his pen name Novalis.

tributed to the first two disciplines, writing on the subjects of thinking and memory, and on the development of concepts in the context of primary and secondary education. In the turn of the twentieth century the ideas of Dörpfeld had acquired prominence within the pedagogical circles. It was widely acknowledged that his writings contained a wealth of deep and original thoughts. These he developed with characteristic clarity in a literary style that was always fitting and lively. He was never content with having gained the acceptance of a colleague or with having refuted an opponent. He wished to reach out for a wider audience and adduce them the fundamental questions that laid at the basis of education. Evidently these questions were too important to be left as the property of a restricted elite only; the educational reform movement of early twentieth century in Germany was a direct verification of this.

Carnap's primary education was naturally shaped according to the ideals of his grandfather. The central principle in Dörpfeld's didactics was that the pupil understand completely whatever is taught, and be given ample opportunity to reflect on what had been learned or discussed. This principle was entertained to enable the pupil to integrate the new material with what he or she already knew from previous studies or experience. Thus, after the death of Carnap's father in 1898, Carnap and his sister were taught at home by their mother Anna. She had thoroughly assimilated the didactic principles of her father and applied them systematically in the education of her children. Everything learned was to be completely understood, digested and connected with other knowledge. This implied that the amount of material taught was restricted to a minimum. The lessons she conducted took usually only about an hour a day, leaving plenty of time for the children to think on their own about the taught material and to 'read around' the subject. This method, as emphasised later by Carnap, contributed essentially towards the development of his faculty of independent judgement and critical thinking: "But she did not simply feed me ready-made answers; her main aim was rather to help me find my own explanations. For example, she might just mention a few facts unknown to me and then say 'The rest of the answer you can now think out for yourself'." In addition to practical didactics, Carnap's mother took pains to explain to the children the ethical attitude of their father which she shared with him. This had a profound effect on Carnap's character formation:

> What convictions, including religious beliefs, anybody had, was for her a morally neutral matter, as long as he would seriously search for the truth and in the forming of his convictions follow his best his insight. This attitude led to a high degree of tolerance . . . I think it was chiefly due to this tolerant attitude of my mother that later, when I abandoned my religious beliefs, I could do so without an internal crisis.[7] [UCLA 1957b, A9–A10]

Whereas Carnap's mother acted as an exemplary model of tolerant thinking in matters of religion, his father, Johann Sebulon Carnap, represented to him 'a man of the people', "cheerful, extroverted, sociable and energetic". In a sense Carnap must have envisaged him as a representative of *Bodenständigkeit* (rootedness in a particular local tradition)

[7]Cited in [Carus (2007), 47].

stemming from the "Bergisches Land", the former Dukedom of Berg, which had become part of Prussia in 1815. Carnap relates the characteristics of the socio-cultural atmosphere in this region as follows:

> The men of this region were known as having a strong sense of external and internal independence. Here the Reformation was not established from above by the authorities, but was carried through by the people themselves. Until 1813 this meant a permanent struggle and resistance against the Catholic regime of the region. Therefore the people themselves had to organise both their church and their school communities and to take care of their own education. For many generations there had been quite a number among the peasants and craftsmen who eagerly read books on religion, philosophy, science, and history, and then gathered in the evening in small groups to discuss their problems. The vivid interest which these people took in their religion and their way of life could sometimes not be satisfied by merely accepting the Lutheran doctrine. The Reformed Church, which was strongly influenced by Calvinist ideas, had many adherents. About the year 1742 some members of this church, among them two ancestors of my father, found the worldly life in the great city of Elberfeld too sinful and intolerable. Eventually a group comprising about fifty families emigrated to the other side of the mountain and founded the town of Ronsdorf, which they called their 'Zion', devoted to a new and better life.[8] [UCLA 1957b, A5–A6]

This remarkable passage is important, not so much for its documentary import as to the background of Carnap's father, but rather as an example of the attitude which Carnap very much considered to be still a part of his personality: "I believe there is still a trace [of this impulse] in me, derived from the strivings of these people for the realisation of a visionary aim, and from their missionary spirit, although [. . .] transferred to secular aims and, in accordance with my more contemplative than active temperament, not expressed in practical activities." [*ibid.*] What is to be stressed here is the primacy of a certain state of mind, or intensity of will, in contradistinction to any particular religious doctrine. What made the impulse stemming from his rural background so forceful was not its specific content but rather the intensity of the feeling and urge for a betterment of the living conditions of the community that accompanied it. As to the abandonment of religious beliefs, Carnap is very open in his Autobiography:

> The transformation of my basic beliefs occurred however not suddenly, but in a gradual development. First the supernatural features in the doctrines of religion disappeared. Christ was regarded not as divine, but as a man among men, distinguished as an important leader in the development of humane morality. Later the idea of God as a personal, though immaterial being, interfering in the course of nature and history in order to reward and punish, was abandoned and replaced by a kind of pantheism. This conception had

[8]Cited in [*ibid.*, 48]

> certain Spinozist features, which came to me less from the works of Spinoza himself than from those of men like Goethe, whose work, personality, and *Lebensweisheit* (wisdom of life) I esteemed very highly. Since my pantheism was thus more influenced by poetical than by philosophical works, it had more an ethical than theoretical nature; that is to say, it was more a matter of the attitude toward the world and fellow human beings than of explicitly formulated doctrines. Later I became more and more convinced that pantheism, if taken not as an emotional-ethical attitude but as a doctrine, could not be scientifically grounded, inasmuch as the events in nature, including those in man and society as a part of nature, can be explained by the scientific method without the need of any idea of God. [Carnap (1963), 7–8]

In 1909 Carnap and his mother moved to Jena where Carnap was to begin his studies in the university. The intellectual atmosphere surrounding Carnap widened considerably. He joined a 'Scientific Club' where students gave talks and discussed contemporary topics of science and philosophy. One of the most frequent topics was Ernst Haeckel's evolutionary naturalism, firmly rooted in the German tradition of positivism. Far from being conventional, he was "admired by some of us, hated by others", as Carnap related. Most importantly, it was here at Jena that Carnap came first in contact with the tradition of ideas stemming from Radical Enlightenment. These took first an outlet in the young Carnap's mind as definite expressions of political ideals. These ultimately turned Carnap into a political activist. To these issues I now turn.

Political activity — *Jugendbewegung* and *Serakreis*

In his early student days Carnap became associated — at first at an intellectual level, then in a more concrete way – with the German Youth Movement [*Jugendbewegung*]. This had its origins in a distinctively German phenomenon, the *Wandervogel* movement, that comprised of social groups of well-off adolescents that rebelled against the inherited and rigid values of their parents and the mass society at large. It drew heavily on German Romanticism and idolized the primitive and medieval societies that were structured on a basis of a feeling of affinity and a conception of simple life. Thus, it was basically a back-to-nature movement whose principal activities were long walks in the German countryside by small groups. During such walks they often sang German folk songs or songs specifically written for the *Wandervögel*. They stayed overnight in barns or with peasants, or, if weather permitted, under the sky. They were emotionally inspired by the heroic ideals and simplicity of bygone ages. Gottfried Gabriel has summed up the motivational background of this movement in the following way:

> The rejection of the 'bourgeois' drugs — alcohol, tobacco, and coffee — has its origin in this movement, and was characteristic of the 'free German students' [*freideutsche Studentenschaft*], who saw themselves as an alternative to the *Burschenschaften* (conservative fraternities) and similar student groups at

> German universities. And its practice of abstinence expresses not an ascetic abnegation, but on the contrary an affirmative *celebration* of the senses, with the goal of a comprehensive reconstruction of life on a 'natural' basis, in the sense of the *Jugendbewegung* (youth movement). [Gabriel (2004) [Awodey & Klein (2004), 8]]

The socialist underpinnings of the leading ideas of this group appealed most strongly to Carnap who witnessed the gradual growth of bureaucracy, spiritual rigidity and intolerance in the public life and cultural atmosphere of the Weimar Republic. In this environment, emotionally strained by the tumultuous situation in the global political arena, the following words of Engels must have rung true for Carnap and his fellow *Wandervögel*:

> The whole sphere of conditions of life which environ man, and which have hitherto ruled man, now comes under the dominion and control of man, who for the first time becomes the real, conscious lord of Nature, because he has now become the master of his own social organization ... Man's own social organization, hitherto confronting him as a necessity imposed by Nature and history, now becomes the result of his free action. The extraneous objective forces that have hitherto governed history pass under the control of man himself. Only from that time will man himself, more and more consciously, make his own history ... It is the ascent of man from the kingdom of necessity to the kingdom of freedom.[9]

Involvement with the group signified a great step for Carnap also at a personal level. He was able to overcome his shyness, gradually emerging from his shell. He had been a somewhat introverted, mathematically inclined student, mainly occupied with theoretical questions but "in Jena he blossomed, found himself not just able to participate as an equal but even to lead." [Carus (2007), 54] Extending his serious involvement with the *Jugendbewegung*, he soon acquired a leading role in a local group working under the auspices of that movement, called the *Serakreis* organised by the publisher Eugen Diederichs. Associated with the *Jugendbewegung* was also an influential movement of social reform, more specifically, a movement attempting to renew the educational system of Germany. Herman Nohl (1879–1960), one of Carnap's philosophy teachers at Jena, had a leading role in these developments. He was able to give the educational reform movement a theoretical basis drawing on Dilthey's *Lebensphilosohie*. Georg Geißler has described the overall significance of Nohl's contributions in this vein as follows:

> Die pädagogischen Reformbestregungen hatten seit der Jahrhundertwende die Wirklichkeit und das Verständnis der Erziehung grundlegen verändert, ein Prozeß, der in Nohls Göttinger Anfängen noch mitten im Gange und mit dem überlieferten Kategorien der Herbartschen Schulpädagogik begrifflich nicht zu erfassen war. Es kam darauf an, diese neue Wirklichleit durch

[9] Friedrich Engels, "Socialism: Utopian and Scientific" in Marx and Engels, *Selected Works*, *II*, 140–1.

> Analyse der Einzelphänomene und deren Einordnung in größere geistige Zusammenhänge verstehbar zu machen und damit in den Griff zu bekommen. Dazu hat Nohls Konzeption einer einheitlichen pädagogischen Bewegung wesentlich beigetragen. Aber eine adäquate Beschreibung und Deutung dieser historischen Erscheinungen konnte nur gelingen, wenn Hand in Hand damit die enstsprechenden systematischen Kategorien entwickelt wurden, und zwar "autonom", aus dem Gegenstandbereich selbst, nich etwa durch Adaptation oder bloßes Überstulpen fremder Begriffssysteme. Weil Nohl immer vom Kern des Erzieherischen aus dachte, führen alle seine Veröffentlichungen auf das Prinzip der pädagogischen Autonomie zurück, was seinem Werk eine große Einheitlichkeit gibt. Den umfassenden Zusammenhang stellt dann die *Theorie der Bildung* dar, der Versuch einer — im Vergleich mit Herbart — von Grund auf neuen pädagogischen Systematik. [Geißler (1979) [Scheuerl, vol. II (1979), 240]]

Carnap held Nohl in high esteem, as we will see later. Indeed, Nohl's conception of *Lebensphilosophie* as well as the integrity of his character constituted a model for Carnap on which he based some of the most general features of his *Weltanschauung*. But as important as these various social phenomena were to the development of Carnap, they were completely outweighed by the catastrophe of the war. Indeed, as to the arousal of political activity among the youth in Germany at that time, the war made all the difference:

> Before the war I, like most of my friends, had been uninterested and ignorant in political matters. We had some general ideals, including a just, harmonious, and rational organization within the nation and among the nations. We realized that the existing political and economical order was not in accord with these ideals, and still less the customary method of settling conflicts of interest among nations by war. Thus the general trend of our political thinking was pacifist, anti-militarist, anti-monarchist, perhaps also socialist. But we did not think much about the problem of how to implement these ideals by practical action. The war suddenly destroyed our illusion that everything was already on the right path of continuous progress. [Carnap (1963), 9]

From then on, the philosophical *habitus* of Carnap comprised a considerable element of a scientific 'engineering attitude', essentially directed at coördinating the activities of society in such a way that they could be removed from the realm of chaotic whim and subordinated to the goal-oriented reason [*der chaotischen Willkür zu entziehen und der zielbewußten Vernunft zu unterwerfen*]. [Carus (2007), 63] This theme will recur frequently in Carnap's philosophical development and it figures prominently in Carnap's mature thought as an essential ingredient of his ideal of explication. But before we can fully appreciate this 'political' element in Carnap's thought, we must investigate the more familiar aspects of his philosophy. At the systematic level, one of Carnap's most important teachers is Frege. I will now proceed to investigate the overall significance of his philosophy to Carnap's thinking.

4.1.2 Frege's influence on Carnap

According to Carnap's own words, the strongest influences upon his philosophical thinking were Frege and Russell. Carnap listened to Frege's lectures in logic at the University of Jena, but was even more influenced by reading his works, mainly after the war. It was not until 1920 that Carnap read Frege's main work, *Grundgesetze der Arithmetik*.[10] From Frege Carnap learned "carefulness and clarity in the analysis of concepts and linguistic expressions, the distinction between expressions and what they stand for, and concerning the latter what he called '*Bedeutung*' (denotation or *nominatum*) and what he called '*Sinn*' (sense or *signification*)". [Carnap (1963), 12] Most importantly, the fundamental tenets of Frege's thought that were to shape Carnap's conception of philosophy throughout his career were the requirement to formulate the rules of inference in logic without any reference to meaning (i.e. the conception of logic as calculus), but also the realization of the significance of meaning analysis (i.e. semantics in a particular sense). Indeed, Carnap summarized his philosophical temperament and sources of interest in the following terms: "I believe that here are the roots of my philosophical interest — on the one hand in logical syntax, and on the other hand in that part of semantics which may be regarded as a theory of meaning". [*ibid.*, 13] I will next briefly adduce the central ideas of Frege's logic that influenced Carnap in his student days. He attended Frege's lecture at the University of Jena during the years 1910–1914. In these lectures Carnap learned about the 'new' logic of quantifiers as well as about the central tenets of logicism the aim of which was to reduce all of mathematics to logic. These ideas were to have massive reverberations in Carnap's later philosophy.

Frege's lectures on logic 1910-1914

The evidence pertaining to the content and style of Frege's lectures on logic at the University of Jena is quite scarce. The only sources of information seem to be Carnap's lecture notes [Reck & Awodey (2004)] and his recollections in his Autobiography, and the reports by Wilhelm Flitner, Gerschom Scholem and Ludwig Wittgenstein. Let us start with Carnap's own recollections of the lectures by Frege:

> Gottlob Frege (1848–1925) was at that time, although past 60, only Professor Extraordinarius (Associate Professor) of mathematics in Jena. His work was practically unknown in Germany; neither mathematicians nor philosophers paid any attention to it. It was obvious that Frege was deeply disappointed and sometimes bitter about this dead silence. No publishing house was willing to bring out his main work, the two volumes of *Grundgesetze der Arithmetik*; he had it printed at his own expense. In addition, there was the disappointment over Russell's discovery of the famous antinomy which occurs both in Frege's system and in Cantor's set theory. I do not remember that he ever discussed in his lectures the problem of this antinomy and the

[10] Two volumes, 1893 and 1903. [Frege (1893/1903)]

question of possible modifications of his system in order to eliminate it. But from the Appendix of the second volume it is clear that he was confident that a satisfactory way for overcoming the difficulty could be found. He did not share the pessimism with respect to the "foundation crisis" of mathematics sometimes expressed by other authors.

In the fall of 1910, I attended Frege's course "Begriffschrift" (conceptual notation, idiography), out of curiosity, not knowing anything either of the man or the subject except for a friend's remark that somebody had found it interesting. {But the very idea of a symbolic notation for concepts seemed attractive to us. Thus we went, and}[11][W]e found a very small number of students there. Frege looked old beyond his years. He was of small stature, rather shy, extremely introverted. He seldom looked at the audience. Ordinarily we saw only his back, while he drew the strange diagrams of his symbolism on the blackboard and explained them. The possibility of a discussion seemed to be out of the question.

Towards the end of the semester Frege indicated that the new logic to which he had introduced us, could serve for the construction of the whole of mathematics. This remark aroused our curiosity. In the summer semester of 1913, ny friend and I decided to attend Frege's course "Begriffschrift II". This time the entire class consisted of the two of us and a retired major of the army who studied some of the new ideas in mathematics as a hobby. It was from the major that I first heard about Cantor's set theory, which no professor had ever mentioned.

[...] In the advanced course on *Begriffschrift*, Frege explained various applications, among them some which are not contained in his publications, e.g., a definition of the continuity of a function, and of the limit of a function, the distinction between ordinary convergence and uniform convergence. All these concepts were expressible with the help of the quantifiers, which appear in his system of logic for the first time. {The last mentioned distinction and some other ones were shown to be based on the difference in the order in which the quantifiers appear, which is, of course, well known today.}[12] He gave also a demonstration of the logical mistake in the ontological proof for the existence of God.

[...] In the summer semester 1914 I attended Frege's course, *Logik in der Mathematik*. Here he examined critically some of the customary conceptions and formulations in mathematics. He deplored the fact that mathematicians did not even seem to aim at the construction of a unified, well-founded *system* of mathematics., and therefore showed lack of interest in foundations. He pointed out a certain looseness in the customary formulation of axioms, definitions, and proofs, even in works of the more prominent mathematicians. As an example he quoted Weierstrass's definition: "A number is a series of

[11]Material inserted in curly braces is from the unpublished version of Carnap's "Intellectual Autobiography". Cited in [Reck & Awodey (2004), 19].

[12]Cited in [Reck & Awodey (2004), 20].

> things of the same kind" ("...eine Reihe gleichartiger Dinge"). {On this he commented with an impish smile: "According to this definition, a railroad train is also a number; this number may then travel from Berlin, pass through Jena, and go on to Munich."}[13] He criticized in particular the lack of attention to certain fundamental distinctions, e.g., the distinction between the symbol and the symbolized, that between a logical concept and a mental image or act, and that between a function and the value of the function. Unfortunately, his admonitions go mostly unheeded even today. [Carnap (1963), 4–6]

In addition to this illuminating report of Carnap we have the documentary evidence of Wilhelm Flitner, a youth movement crony of Carnap and a seminal figure in inaugurating the foundation of the Jena *Volkshochschule*, a college for adult and further education, in 1919. Flitner ultimately came a professor and an influential educational thinker. His reminiscences of Frege's lectures bring further clarification to Carnap's relation with Frege, and the influence of Frege on the development of Carnap's own thinking:

> [Carnap's] interest in logical problems led him to the lectures of an associate professor [Außenordentliche Professor] who was then almost completely unknown: Gottlob Frege. Ernst Abbe[14] had recommended Frege to the faculty [to become associate professor] in 1883, pointing to Frege's excellent logical investigations; Abbe's successor Siegfried Czapski had regarded him — as we know from his children — as the most important thinker at the University of Jena.
>
> Nevertheless, Frege struggled to even have his lectures take place. "Tres faciunt collegium" ["Three makes a collegium"] was then the rule, with the lecturer counting as a third person. For several semesters the lecture would have had to be cancelled had Carnap not found a second hearer. Therefore he enlisted me, and I attended the extraordinarily fascinating logical lecture course on "Begriffschrift"; and the following semester Julius Frankenberger's brother, who majored in mathematics and physics, played the role of the third person. For Carnap, these lectures became the foundation of his later philosophy; he saw Frege's achievements as the most important developments in logic since Aristotle and Leibniz. [...]
>
> During his time as a student, Carnap never spoke a word with Frege; likewise, I only exchanged a few insignificant words with him, at his door, when I had to deliver something. Putside of the university one didn't dare to address Frege. In spite of the fact that he was our neighbor on the Forstweg, one rarely saw the not very tall man, except when he walked over the Forstweg bridge, looking downwards and a hand on his back, and then disappearing in his house. In his lectures he rarely glanced at his students; he was exclusively

[13] Cited in [Reck & Awodey (2004), 21].

[14] Physicist and professor at the University of Jena, and the founder and owner of the renowned Zeiss optical works which was famous for its educational program. This program was a kind of an implementation of the *Bildungsideal*, providing the employees with opportunities for mental and physical self-improvement.

concerned with symbols he wrote on the board and explained in a totally incontroverted manner, thus wholly focused on the subject matter of "logic". [Flitner (1986), 126–127][15]

These quotations provide a more balanced and richer account of Frege as a teacher, and his relation with students, than is usually mediated in the literature. At the very least they demonstrate the enthusiasm and devotion with which Carnap attended the lectures. The surviving notes that Carnap took in those lectures have subsequently been edited by Erich Reck and Steve Awodey, and translated into English [Reck & Awodey (2004)].[16] The notes were originally written systematically in the Stolze-Schrey shorthand that Carnap used in his student years. Of most interest from the point of view of the subsequent development of Carnap's own philosophy are the sections devoted to concepts in "Begriffschrift I" and the general remarks about definitions, thought content of sentences, sense and reference, and functions. The treatment of all these notions was thought-provoking and profound. Surprisingly, Carnap failed to recall the essentials of Frege's insights pertaining to concept symbols and their referents, and erroneously attributed in *Meaning and Necessity* to Frege the view according to which the extension of a concept symbol is the set of objects falling under it.[17] This contradicts the very idea documented by Carnap himself in his notes on Frege's lectures. [Frege (1910–1914) [Reck & Awodey (2004), 74]] To make our account of the influence of Frege on Carnap more concrete, it is expedient to survey briefly the contents of the lectures and the specific notation used in them.

Frege based his exposition in the 1901–1914 lectures on the general framework contained in his *Grundgesetze der Arithmetik*. In this mature work the logical system had already been crystallized and therefore the mode of expression did not vary much in the Jena lectures, as is evidenced by Carnap's notes. The vocabulary of Frege's logical system

[15] Translation in [Reck & Awodey (2004), 22].

[16] An earlier German edition of the notes of the first two lectures is [Frege (1996)], edited by Gottfried Gabriel.

[17] "Our pair of concepts [Extension and Intension] is, like Frege's, intended to serve for the purposes of semantical meaning analysis. Our two concepts may be regarded, like Frege's, as representing two components of meaning (in a wide sense). The concepts of sense and of intension refer to meaning in a strict sense, as that which is grasped when we understand an expression without knowing the facts; the concepts of nominatum and extension refer to the application of the expression, depending upon facts.

A decisive difference between our method and Frege's consists in the fact that our concepts, in distinction to Frege's, are independent of the context. An expression in a well-constructed language system always has the same extension and the same intention; but in some contexts it has its ordinary nominatum and its ordinary sense, in other contexts its oblique nominatum and its oblique sense.

Let us, first, compare the extension of an expression with its ordinary nominatum; it seems that these concepts coincide. With respect to predicators, Frege does not seem to have explained how his concepts are to be applied; however, *I think that Church [. . .] is in accord with Frege's intentions when he regards a class as the (ordinary) nominatum of a predicator (of degree one) — for instance, a common noun — and a property as its (ordinary) sense.* As an example, Church states that the nominatum of 'unicorn' is the null class, and its sense is the property of unicorn-hood. And here the extension is likewise the class in question. With respect to a sentence, its truth-value is both the ordinary nominatum and the extension. And in the case of an individual expression the ordinary nominatum and the extension is the individual in question. Thus we have this result:

[. . .] For any expression, its ordinary nominatum (in Frege's method) is the same as its extension (in our method)." [Carnap (1947 [1956]), 125] My emphasis.

is very much akin to the usual machinery of predicate logic. Indeed, it comprises the following classes of symbols:

- Propositional letters: $P, Q, \ldots$
- Individual letters; $a, b, \ldots$
- Predicate letters: $F, G, \ldots$
- Higher-order predicate letters: $\alpha, \beta, \ldots$
- Function letters: $f, g, \ldots$
- Variables: $x, y, \ldots, X, Y, \ldots$

The last class consists of special letters for variables involved in quantification. Furthermore, atomic formulas comprise the following:

P	for a propositional letter P
$a = b$	for individual letters a, b
$F(a, b, \ldots)$	for a predicate letter F and for individual letters $a, b, \ldots$
$\alpha(F, G, \ldots)$	for a higher-order predicate letter α and predicate letter $F, G, \ldots$

A peculiar feature of Frege's system is the 'horizontal' function

$$—\varphi$$

which Frege occasionally calls the "content stroke" and that can be applied to *any* argument φ (proposition or not). It yields the value *true* if and only if the argument is true, and *false* otherwise. Hence, we have, e.g.

$$—(a = a) = \top$$

and

$$—\left[\frac{\sqrt{5}}{1+\sqrt[5]{5^{\frac{3}{4}}\left(\frac{\sqrt{5}-1}{2}\right)^{\frac{5}{2}}}-1} - \frac{\sqrt{5}+1}{2}\right] e^{2\pi/\sqrt{5}} = \perp,$$

where "$\top$" denotes "true" and "$\perp$" denotes "false", respectively.

Formulas are built from the atomic formulas applying the horizontal function and the propositional connectives *negation* and *conditional*:

$\neg\varphi$	negation
$\varphi \to \psi$	conditional

and the universal quantifiers:

$\forall x \varphi(x)$	for an individual variable x
$\forall X \varphi(X)$	for a predicate variable X

The other familiar connectives and quantifiers are also often applied:

$\varphi \& \psi$	conjunction
$\exists x \varphi(x)$	existential quantification

The axioms of the *Begriffschrift I* are based principally on the axioms of *Grundgesetze*, but form actually a sort of a hybrid of the axioms presented there and in Frege's other published works. Thus, the first three axioms of the *Begriffschrift I* read:

Axiom I: $P \to P$
$P \to (Q \to P)$

Axiom II: $\forall x F(x) \to F(a)$
$\forall X \alpha(X) \to \alpha(F)$

Axiom III: $g(a = b) \to g(\forall X(X(a) \to X(b)))$

The last axiom (Axiom III) yields *Leibniz's law* as a special case:

$$a = b \to \forall X(X(a) \to X(b))$$

which is also known as the law of "indiscernibility of identicals" (not "identity of indiscernibles" as is sometimes claimed; this is the converse of the former, namely "$\forall X(X(a) \to X(b)) \to a = b$"). The functional formulation of the Axiom in Frege's system permits its application in *any* context. It is also known as *Frege's principle*: If the expression B may be derived from expression A by substituting an instance of some symbol contained in A by another symbol that has the same *extension*, then A and B have the same extension. In contradistinction to the *Grundgesetze*, Frege incorporated only three axioms to his exposition in the Jena lectures. He had dropped Axioms IV, V, and VI that played a prominent role in the *Grundgesetze*. These axioms were:

Axiom IV: $\neg(P = \neg Q) \to P = Q$

Axiom V: $\{x : \varphi\} = \{x : \psi\} \leftrightarrow \forall x(\varphi \leftrightarrow \psi)$

Axiom VI: $a = (\iota x.x = a)$

The first is used in *Grundgesetze* (§51) to prove "propositional extensionality", i.e., $(P \leftrightarrow Q) \to P = Q$. Russell's well-known antinomy arises quite directly from the Axiom V (governing the terms representing the extensions of concepts represented by "φ"). In the Axiom VI Frege introduced a description operator "$\iota x.\varphi$" denoting the unique individual, if there is one, satisfying the condition expressed by "φ", effectively explicating the idea of "the x such that φ". The system without these axioms, which formed the basis of Frege's lectures on the *Begriffschrift*, may be described as the 'inferential part' of the system of *Grundgesetze* without the constructive machinery for building up logical objects. [Reck & Awodey (2004), 34]

The rules of inference fall into two categories: (i) rules involving a single formula as a premise, and (ii) rules applying two formulas to draw a conclusion. In the first category are rules such as:

$$\text{Exchange}\ \frac{P \to Q \to R}{Q \to P \to R}$$

which applies equally for any reordering of any number of constituent propositions $P, Q, R, \ldots,$

$$\text{Transposition}\ \frac{P \to Q \to R}{P \to \neg R \to \neg Q}$$

which also applies equally for any other conditional position. The rule for 'redundant' iteration of the conditional premise is called as the rule of "Collapsing":

$$\text{Collapsing}\ \frac{P \to P \to Q}{P \to Q}$$

which holds similarly in the presence of additional conditions (in any position). The rule for generalization is:

$$\text{Generalization}\ \frac{\varphi(a)}{\forall x \varphi(x)}$$

where x may not already occur in $\varphi(a)$. This rule holds analogously for function and predicate letters, as in

$$\text{Generalization(Pred)}\ \frac{\varphi(F)}{\forall X \varphi(X)}$$

where X may not already occur in $\varphi(F)$. The rules pertaining to inference involving two premises are:

$$\text{Detachment}\ \frac{P \qquad P \to Q}{Q}$$

$$\text{Cut}\ \frac{P \to Q \qquad Q \to R}{P \to R}$$

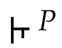
and

$$\text{Negation}\ \frac{P \to Q \to R \qquad P \to \neg Q \to R}{P \to R}$$

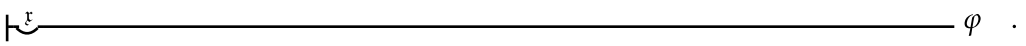
All these rules can be applied also in the case when there are several conditions present. This, then, is the skeleton of Frege's system. I will now proceed to adduce some of the important results that are demonstrated within it.

It is expedient to give a brief survey of the symbolism of the *Begriffschrift*. The basic symbols by means of which the 'well-formed' formulas are constructed are the symbol for a conditional statement "if P then Q":

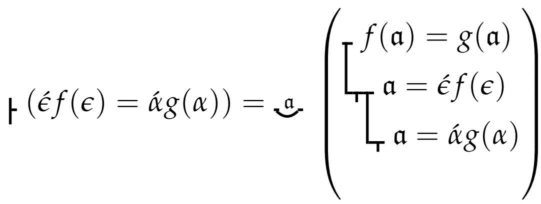
the symbol for negation "$\neg P$":



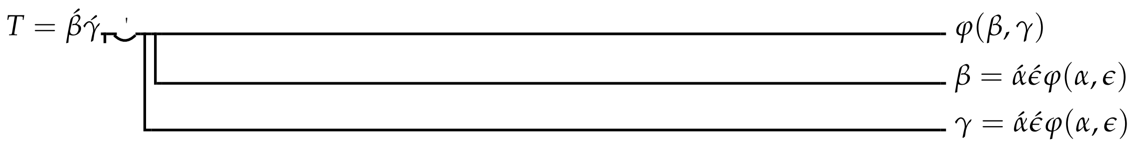
and the symbol for universal quantification "$\forall x \varphi$":



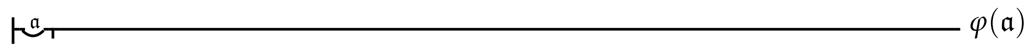
To illustrate the expressive power of the symbolism, consider the following example from Frege's correspondence with Russell giving an expression for the Basic Law V of the *Grundgesteze*:



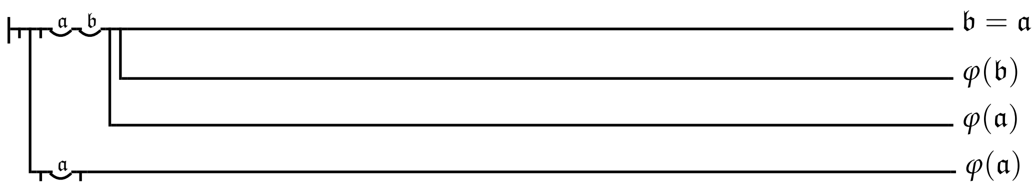
or another expression from Russell's correspondence with Frege:



Frege's system allowed to define numerical statements about concepts (of the number of objects falling under them) in the following manner. First, to express the statement *The number of objects falling under the concept is **0*** he used the following notation:



Proceeding in a similar manner, the numerical statement *The number of objects falling under the concept is **1*** could be formulated as:



As Carnap wrote in his lecture notes, "[t]he numerical statement concerns the kind of satisfaction; it is a 2^{nd} level concept indicating the features of a concept." [Reck & Awodey (2004), 84] This conception about statements of number was adopted by Carnap in his later work [Carnap (1930a)], where he effectively defined the notion "the number of the concept f is two" as "There is an x and there is a y such that x is not identical with y, x falls under f, y falls under f, and for every z it is the case that if z falls under f, z is identical with x or with y". [Carnap (1930a [1959]), 21/141][18] But what goes unmentioned is Frege's reservation that this definition (and definitions analogous to it) do not define the numbers themselves (i.e. **0**, **1**, **2**, etc.), but merely the phrases "the number N belongs to [the concept F]". Contextual definitions of this kind do not allow us to distinguish between individual numbers, such as **0** and **1**, for example. Carnap only later noticed the difference between his position and Frege's, and made remarks about it accordingly. In *Meaning and Necessity* Carnap elaborates on this point:

> [Frege takes as his definiens not 'the property Equinumerous to f', but 'the extension of the property Equinumerous to f' which means the same as 'the class equinumerous to f'. Now it is interesting to see that Frege adds to this definition a footnote [...] which says: "I believe that instead of 'extension of the property' we might say simply 'property.' But two objections would be raised: ...I am of the opinion that both of these objections could be removed; but that might lead here too far." Thus Frege considers here the simpler procedure which we now adopt. He seems to regard it as feasible but does not pursue it any further. In his later work he again defines cardinal number in the way stated above, without even mentioning an alternative possibility. His chief reason for regarding cardinal numbers as classes of properties rather than as properties of properties seems to be his view that cardinal numbers are independent entities, in combination with his general conception that classes are independent entities, while properties are not. However, I find his reasoning on this question not quite clear and far from convincing.] [Carnap (1947 [1956]), 116]

[18] The German original reads: "[...] an der Zahl Zwei als Kardinalzahl, d. h. als Anzahl eines Begriffes. Wir definieren: 'die Anzahl des Begriffes f ist zwei' soll bedeuten 'es gibt ein x und es gibt ein y derart, daß x nicht identisch mit y ist, x unter f fällt, y unter f fällt, und daß für jedes z gilt: wenn z unter f fällt, so ist z mit x identisch oder mit y identisch' ". [*ibid.*]

As Gottfried Gabriel notes, this implies that Carnap does not subscribe to Frege's categorial distinction between "complete [*abgeschlossenen*]" objects and "unsaturated [*ungesättigen*]" concepts. [Gabriel (2007) [Friedman & Creath (2007), 68]] It has been suggested that Carnap might have misunderstood Frege's distinction. This is rendered plausible by the paragraph §33 in the *Aufbau* where in which Carnap elucidates the notion of "unsaturated" symbols. After stipulating that classes are quasi-objects since they are extensions, and that the class symbols do not therefore have independent meaning, Carnap goes on to discuss Frege's position with respect to classes: "Frege has already shown that extension symbols, and thus the class symbols, are *incomplete symbols*. [...] According to Russell, it is irrelevant for logic whether or not there are actual objects which are designated by class symbols, since classes are not defined by themselves, but only in the context of total sentences ('no class theory')." [Carnap (1928 [1967]), §33] The misunderstanding or misinterpretation on Carnap's side is effectively concealed in the English edition of *Aufbau* by an unfortunate translation of "*ungesättig*" as "incomplete". The term "incomplete" is used here by Carnap in the sense in which Russell applies it in connection with class symbols, i.e., symbols that do not have an independent meaning in themselves. Following Russell we can certainly say that expressions for extensions of concepts are, in the sense intended, *incomplete* symbols, but Frege maintained that they are not *unsaturated*, because they designate *complete objects*. [Gabriel (2007) [Friedman & Creath (2007), 69]] Hence, there does seem to be some evidence for thinking that Carnap failed to grasp Frege's distinction in his later writings. This is, of course, congruent with the misrepresentation by Carnap of Frege's notion of the extension of concept words that we alluded to earlier.

"Making sense" — What did Carnap carry over from Frege?

One important feature that Carnap adopted from the exposition of Frege was the notion of logic as a calculus. Although not at all congruent with Frege's underlying metalogical convictions (the idea of an uninterpreted calculus was repugnant to Frege), this idea may nonetheless have been instigated by Frege's lectures because, as we have seen, Frege had dropped the last three axioms of his *Grundgesetze* system out of the logic lectures, leaving only a skeleton with purely inferential character, with no constructive machinery. This might also have given Carnap the impression (which he reported in his Autobiography) that Frege had circumvented the difficulties brought up by Russell's antinomy (the famous "set of all sets that are not elements of themselves").[19] Since Frege did not mention the paradox in his lectures, Carnap was inclined to think that it was no more a problem for Frege's system. The true reason for it having been left unmentioned is that the exposition of the restricted 'skeleton' system did not give rise to questions pertaining to the logicist program, and hence to the paradox. The resulting system had no domain of its own but provided inferential apparatus for the study of reasoning in other domains. Carnap crystallized this idea in 1932, possibly recalling the contents of Frege's lectures:

[19]Also known as **Russell's paradox**: There is no set $S = \{X : X \notin X\}$. The axiom schema of comprehension fails for the formula $\varphi(X)$ interpreted as "$X \notin X$". This is easily seen, if we consider the schema $\forall Y : Y \in S \iff Y \notin Y$. Now, substituting S for Y we have $S \in S \iff S \notin S$, which is clearly impossible.

> The propositions of logic and mathematics [...] are of great significance for science, since they aid the transformation of [scientific] propositions. [...] Logic and mathematics are not sciences with a domain of objects of their own. [...] The assumption of "formal" or "ideal" objects, as opposed to the "real" objects of the empirical sciences, is dropped. [Carnap (1932a), 433][20]

Carnap, as is well known, held such a view for a relatively long time, at least until the beginning of his 'semantical' period in the mid-1930s. It is difficult to determine the exact point at which he came to re-evaluate his conception of logicism. On the one hand, he can be said to have retained a broadly logicist outlook on the foundations of mathematics throughout his career, but never in the extreme sense of formalism (as propounded by, for example, Professor J. Thomae, Frege's colleague at Jena, with whom Frege had a polemical exchange in the *Jahresbericht der deutschen Matematikervereinigung* in 1906–1908). Carnap's logicism was more in the spirit of Hilbert, as his later writings on the foundations of mathematics attest.[21] Carnap's adoption of the semantical viewpoint – a viewpoint taken only later by Alfred Tarski[22] who first joined Quine in criticizing Carnap's approach – radically transformed his conception about the content of logicism. This transformation resulted in a more nuanced and richer conception of mathematics which squared with Carnap's fundamental distinction between theoretical and observational concepts that remained important for him throughout the latter part of his career.

Another feature in Carnap's thinking that was profoundly influenced by Frege was his treatment of higher-order logic with simple types (as opposed to the ramified types) in the late 1920s and early 1930s.[23] Indeed, this was an area in which Carnap was to have a significant role in guiding the development of modern logic in general. It has been argued that in the light of the relevant evidence, Carnap's textbook *Abriss der Logistik*, published in 1929, but completed and circulated as early as 1927, seems to be the first systematic treatment of higher-order logic with simple types. [Reck & Awodey (2004), 39] It has also been remarked that the frequently cited sources of higher-order logic, Leon Chwistek's "The Theory of Constructive Types I" (1924) and Frank Ramsey's "Foundations of Mathematics" (1925) contain much less detail than was already present in Frege's works. Most importantly, Carnap's *Abriss* and other related works pertaining to the theory of types were known to Gödel whose revolutionary paper "Über formal unentscheidbare

[20]Translation by Reck & Awodey.

[21]One of the most illuminating writings in this respect is [Carnap (1966b)].

[22]Tarski first developed his variant of model theory as a part of the metatheory of algebra.

[23]This is not a proper place to provide a technical exposition of the theory of types, but I will say something in the way of conveying the gist of the theory — as well as the notion of type — and its implications, especially as concerns its effect on the structuring of the universe into types. Russell defines type as "the range of significance of a propositional function, that is, as the collection of arguments for which the said function has values." [Russell (1908), 163] Climbing up the definitional hierarchy, we begin with the lowest type which is simply the class of (all) individuals. In [(1908)] Russell characterizes individuals negatively, stating them as being devoid of logical complexity and hence as different from propositions and propositional functions. Incidentally, already this important, albeit simple requirement is tantamount to an exclusion of the possibility that quantification over individuals might already involve a vicious circle. Type 1 contains all the (definable) classes of individuals; type 2 all the (definable) classes of classes of individuals; and so on. This iterative procedure constitutes the essence of the simple theory of types.

Sätze der *Principia Mathematica* und verwandter Systeme I" [Gödel (1931)] is generally regarded as one of the most important sources of the whole field of modern logical theory. In fact, Carnap devoted about ten odd pages to the Theory of Types in the *Abriss* which comprised roughly 110 pages in total. This proportion of material pertaining to type theory in a book that was intended as an *elementary* exposition of logic and its applications, testifies of its importance in the mind of the author. Indeed, as Reck and Awodey have remarked: "under Frege's influence, simply-typed higher-order logic was never just a device for avoiding contradiction for Carnap, but was in the very nature of logic, lending it an inherent plausibility that other conceptions lacked." [*ibid.*] In sum, the particular way in which Frege had presented his conception of logic in the lecture courses "Begriffschrift I" and "Begriffschrift II", made it quite natural for Carnap to adopt two stances that were unusual at the time: (i) he worked with a higher-order logic based on simple types and (ii) he used higher-order logic as an *inferential* framework. [Reck (2007) [Friedman & Creath (2007), 181]]

The feature in Carnap's thinking that was perhaps most intensively permeated by Fregean conceptions was his general account of the domain of semantics. This feature received its most balanced treatment in Carnap's *Meaning and Necessity*. We have already pointed out some issues that were explicitly discussed by Carnap in connection with Frege's ideas. We will return to some of these (along with other issues) in Chapter 5. It might be pertinent to complete the discussion on Frege with the following statement of Carnap about the overall significance of Frege on his thinking:

> Furthermore, the following conception, which derives essentially from Frege, seemed to me of paramount importance: It is the task of logic and mathematics within the total system of knowledge to supply the forms of concepts, statements, and inferences, forms which are then applicable everywhere, hence also to non-logical knowledge. It follows from these considerations that the nature of logic and mathematics can be clearly understood only if close attention is given to their application in non-logical fields, especially in empirical science. [Carnap (1963), 12]

But as much as Carnap was influenced by Frege, there were four fundamental differences between them: (1) Carnap's epiricism, (2) Carnap's view that formal reasoning in logic (and thus in arithmetic) can be detached from its content, (3) Carnap's inclusion of geometry in logicism, following Russell, and (4) Carnap's dismissal of ontological concerns as pseudo-questions. [Gabriel (2007) [Friedman & Creath (2007), 77]] Carnap's deterioration from the viewpoint of his teacher was, however, quite a complicated process, taking several years. I will now turn to adduce other influences that proved crucial for Carnap's intellectual orientation during his final years at Jena.

4.1.3 A system of knowledge — science and logic

As a result of his broad and penetrating studies, the thoroughgoing gulf between philosophy and sciences had aroused Carnap's intellectual uneasiness. Although he had already in 1916, in a letter to a friend, expressed his dissatisfaction with the level of sophistication of philosophical writing on science, relating that "Even a mind as sound as Mach's frequently makes me shake my head" [ASP/WF (1916c)],[24] by 1920 he had become much more aware of the depressing situation. The discrepancy had come about largely because of the very fast development of the sciences that proceeded mathematically (sciences using mathematical tools, methods and models in their explanatory tasks) and because these sciences were themselves incapable of appraising critically their conceptual foundations and the methods they utilized. Carnap maintained that both the sciences and philosophy were to blame for this state of affairs:

> Part of the blame rests with philosophy, which has often failed to comprehend the viewpoints of those rapidly developing sciences; on the other hand, part of the blame goes to the sciences, which were occupied more with conquering new territories than with securing and carefully integrating what had been gained. In short, through the fault of both sides a mutual alienation developed. [ASP (1920d), 1–2][25]

He experienced very intensely the differences and discords between the two, but came to appreciate that the discrepancy was not entirely due to either separately, but was symptomatic of the situation. Carnap makes this clear in a chain letter to the members of the *Sera Kreis*, utilizing military metaphors as an expressive device (an understandable manner of writing at the time):

> What happened was unavoidable given the rapidity of an advance in mobile warfare: communications broke down between front and staff. To the philosophers at headquarters, in the absence of bulletins [from the front], the situation became less and less clear; it is especially unfortunate that although Kant was still alive at the time of the first work on the fundamentally important non-Euclidean geometries, he received no bulletins about them — which [if he had received them] might have been of great benefit both to his system and to those geometries. At the front, the lack of communication could have been even more damaging if there hadn't fortuitously been a few leaders who made up for their lack of strategic schooling with native insight and clear-sightedness: thanks to men like Gauss and Helmholtz as well as Riemann and Hertz, among many others, who combined outstanding talent in their specialties with great system-building powers and sureness of instinct for the big picture, by and large the right paths were followed. [ASP (1920), 2][26]

[24]Cited in [Carus (2007), 91].
[25]*ibid.*
[26]*ibid.*, 91–92.

However, the singular cases of individuals of such calibre did not exemplify the prevailing intellectual situation, especially not at the turn of the twentieth century. The process of gradual alienation of science from philosophy was a fact, and it seemed to accelerate with increasing rate. Carnap reacted to this acutely:

> For some time now in geometry, more recently also in arithmetic and analysis, and now in physics as well, it is becoming clear that at certain point the bulletins from the front don't at all fit with the official map at the headquarters. After making do for a while with saying, at the headquarters, 'they don't themselves seem to know where they are!', and at the front, 'what do we care about the outdated map back there!', the insight is dawning that neither accusation is altogether unjustified: the exact sciences often work with concepts (which in some cases turn out to be their most important ones) whose meaning they can't give precisely; and on the other hand the traditional methods of philosophy can't help much with this. [ASP (1920d), 2–3][27]

Some scholars had already responded to the drastic and revolutionary situation Carnap here alludes to. These scholars were working mainly under the disciplines of philosophy, mathematics and physics. Their aim was to draw a new 'comprehensive map' that would equally satisfy all the parties on the sides of both philosophy and science. Moreover, the map was required to give an accurate image of the real state of knowledge, constituting in effect a "system of knowledge" (*System der Wissenschaft*) that would accommodate all the relevant theoretical considerations under its purview. Carnap very much wished to be part of this attempt, maintaining that such a system should be constructed with "logically consistent foundations and systematic construction of concepts", a system that is "capable of comprising all the insights of the special sciences and of presenting them with the greatest possible simplicity and unity". [ASP (1920d), 3][28]

4.2 The Neo-Kantian Roots of Carnap's Thought

> For the transcendental idealist, the object of knowledge is . . . neither immanent nor transcendentally "given" ["*gegeben*"], but rather "posed as a problem" ["aufgegeben"]. — Rickert (1921)

In the preceding section we examined the evidential support for the influence of Frege on Carnap, and found it ample. The high esteem in which Carnap held his old master reflected an emotional attachment and intellectual respect that remained characteristic attitudes throughout his life. Undoubtedly Frege was one of the brightest stars in Carnap's philosophical firmament. But the appraisals of Frege hide the indisputable relevance

[27] *ibid.*, 92.
[28] *ibid.*

of other sources of influence that played at least an equally important role in Carnap's philosophical development and the process of formulating his central ideas. These 'hidden' influences derive from neo-Kantianism, the foundational issues related to geometry, and perhaps most surprisingly, Husserl's philosophy. They are clearly visible in Carnap's early work and their apparent disappearance in Carnap's later work is actually only a consequence of a choice of a new linguistic framework by Carnap that does not explicitly use the concepts of the earlier constitutional program. This constitutional program consisted of a radically new insight into the epistemological questions posed by the neo-Kantian tradition. In part, it was also a deeply-reflected response to the challenge to realize the vision of an logicist-empiricist program of philosophy *in concreto* as inaugurated by Russell in his *Our Knowledge of the External World*.

4.2.1 The Kantian legacy

Kant's project in the *Critique of Pure Reason* (*Kritik der Reinen Vernunft*) has been interpreted, roughly speaking, along two diametrically opposite lines. The first one takes the suggestion of Kant about his "Copernican revolution" seriously. The second one interprets it as an instigation to return to an anthropomorphic view of the nature of reality and cognition, in keeping with the romantic tradition of *Naturpilosophie*. The first line of interpretation can be termed "Copernican", the second one "Anti-Copernican". But why the label "Copernican" in the first place? This simile is based on the observation that just as Copernicus had explained the movements of the stars and planets by suggesting that their apparent movements are partly due to the state of motion of the observer, so Kant proposes to explain the application of the *a priori* principles of the mind to objects by suggesting that the appearance of objects is due to the mind's innate constitution and in so far as the categories of the mind regulate observation, "objects conform to the mind". This is Kant's 'critical idealism'. It is motivated by the idea that there is a sort of half-way house between the realism of an *intellectus ectypus* or passive mind and the idealism of an *intellectus archetypus* or creative mind. The essential content of such a characterization is made clearer if we consider the status of natural sciences in Kant's system. Kant's main contention is that the fundamental task of natural sciences is to anticipate what we shall experience. Every scientific judgement with claims of truth, when analysed in full, is essentially a statement that under such and such circumstances, we will have such and such experiences. In short, in our scientific judgements we are always making statements about our possible experience (in the sense of both prediction and retrodiction). In the first edition of the *Critique of Pure Reason* Kant did not intend to *presuppose* that our knowledge in the domain of mathematics or metaphysics is synthetic *a priori*, but instead proceeded to investigate generally the basic elements of *any* experience, and then showed from the results of these investigations that we *de facto* have *a priori* cognition, not only in metaphysics, but also in mathematics and even physical science. The general objective of his "Deduction of the Pure Concepts of the Understanding" is *both* "to demonstrate and make comprehensible the objective validity of its concepts *a priori*" [A xvi], i.e. to both *prove* that we have synthetic *a priori* cognition in mathematics, science and metaphysics and then explain *how* such knowledge is possible. Kant's method in tackling this double task is to try a procedure analogous to the "first thoughts of Copernicus" [B xvi] — hence the label "Copernican revolution". In Kant's own words, just as Copernicus,

> when he did not make good progress in the explanation of the celestial motions if he assumed that the entire celestial host revolves around the observer, tried to see if he might not have greater success if he made the observer revolve and left the stars at rest,

so

> in metaphysics we can try in a similar way regarding the intuition of objects. If intuition has to conform to the constitution of the objects, then I do not see how we can know anything of them *a priori*; but if the object (as an object of the senses) has to confirm to the constitution of our faculty of intuition, then I can very well represent this possibility to myself. Yet because I cannot stop with these intuitions, if they are to become cognitions, but must refer them as representations to something as their object and determine this object through them, I can assume that the concepts through which I bring about this determination also conform to the objects, and then I am once again in the same difficulty about how I could know anything about them *a priori*, or else I assume that these objects, or what is the same thing, the experience in which alone they can be cognized (as given objects) conforms to those concepts, in which case I immediately see an easier way out of the difficulty. [B xvii]

Hence, the assumption that we can find fundamental conditions of the possibility of our own experience to which the objects of our experience must conform, is the basis for Kant's first claim of autonomy. [Guyer (2006), 50] It comprises the idea that sensibility and understanding, as two main faculties of the mind, contain "the constitutive principles *a priori* for the faculty of cognition (the theoretical cognition of nature)." [CPJ, 5:196] In the *Critique* Kant states his position explicitly, making it clear that the claim for autonomy must be taken in a very strong sense: "[...] as exaggerated and contradictory as it may sound to say that the understanding is itself the source of the laws of nature [...] such an assertion is nevertheless correct and appropriate to the object, namely experience." [A 127] It is precisely this tenet that instigated a vehement philosophical debate in the centuries to come, and that resulted, roughly, in the splintering of the Kantian influence into two rival schools, as I have hinted at above. These developments are far too entangled to be delved into in the context of this dissertation.The secondary literature on the reception of Kant's thought is vast. I settle here for providing a list of the treatises that I have found most illuminating. [...] I will investigate, in some detail, the relation of Kantianism to the phenomenon of *Lebensphilosophie* in Germany in the next section. Moreover, I will hint at the importance of neo-Kantian influences in Carnap's struggle with the philosophical problems of space and time in his early work in the sections that follow. But overall, I have very little to say about the *general* significance of neo-Kantianism and its place in the philosophical tradition of the late nineteenth century and early twentieth century. An interested reader should consult the superb treatises mentioned in the footnotes.

4.2.2 *Lebensphilosophie* and Kantianism

In Jena, before the war, one of Carnap's most influential teachers had been Herman Nohl, an educator and a *Privatdozent* ('a private lecturer') who was a student of Wilhelm Dilthey. He represented in a pure form what could be called *Lebensphilosophie*, a 'school', or rather an intellectual orientation within philosophy, that went against neo-Kantianism that was prevalent in the German universities at the time. Carnap made very kind remarks about Nohl in his *Autbiography* which hint at the significance that this humble philanthropist had for Carnap and his philosophical orientation:

> I remember with special pleasure and gratitude the seminars of Herman Nohl (at that time a young instructor in Jena), in philosophy, education, and psychology, even when the topic, for example, Hegel's *Rechtsphilosophie*, was often somewhat remote from my main interests. My friends and I were particularly attracted by Nohl because he took a personal interest in the lives and thoughts of his students, in contrast to most of the professors in Germany at that time, and because in his seminars and in private talks he tried to give us a deeper understanding of philosophers on the basis of their attitude toward life [*Lebensgefühl*] and their cultural background. [Carnap (1963), 4]

The essential import of Nohl's thinking resided in its ability to give the prevailing attitudes behind the new social movements (foremost the German *Jugendbewegung* and educational reform movements) a theoretical basis. This basis was built on the central ideas of Dilthey's *Lebensphilosophie* the most important ingredient of which was the insistence on the primacy of life even in the abstract realm of philosophical theories. [Gabriel (2004) [Awodey & Klein (2004), 9]] Dilthey's doctrine of *Weltanschauung* was the conceptual axis around which these aspirations winded. Carnap was very much influenced by these ideas – a major factor arousing his interest in them being without doubt his intensive engagement with the student movements alluded above. Carnap did not, however, rest content with the formulations of Dilthey and Nohl. He attempted to sharpen the conceptual ingredients behind such general philosophy and crystallized them in a form that suited better his own interests and aspirations. He made a distinction that characterized the components of any given "world picture" [*Weltbild*]. He analysed the content of any such world picture into elements that expressed values or sentiments [*Lebensgefühl*] on the one hand, and elements that were purely cognitive, effectively expressing beliefs [*Anschauungen*], on the other. The latter was coincident with what Kant had called the domain of (theoretical) judgement, characterized further with the property that its elements could be either true or false. The former domain of *Lebensgefühl* was outside meaningful linguistic expression, not being amenable to rational judgement. However, Carnap's innermost conviction was that better and more reliable knowledge could fundamentally transform the domain of *Lebensgefühl*. Hence, the basis of all his philosophical work was built around the realm of knowledge, focusing on the task of devising better and more refined concepts, more adequate theoretical constructions, and methods of testing such theories. This general theme was to remain a fundamental characteristic of all his work throughout his long career. The contrast between the two domains that Carnap had distinguished was exemplifying in his mind the fundamental cleavage that exists in our knowledge acquisition procedures and practices between the intersubjectively describable domain of constructed theoretical knowledge and the subjective domain of experiential content. Hence, according to Carnap, Goethe's critique of Newton was misguided in the sense that it illegitimately imposed the domain of *Lebensgefühl* on the realm of science, since in physics the aspects of experience that were the focus of Goethe's investigations are irrelevant from the point of view of what is "first according to nature", πρότερον τῃ φύσει (*proteron te phusei*), and thus the former (phenomena) consisting of colours, sound qualities, tactile sensations, etc. are rather always an issue about being first to us, πρότερον πρός ἡμᾶς (*proteron pros ēmas*). To conflate these domains (*proteron te phusei* and *proteron pros ēmas*) in the order of nature *à la* Goethe was to instill a dangerously great authority to *Lebensgefühl*, thereby distorting our picture about what could be rationally judged. For Carnap the task of science was essentially to liberate us from the modalities and inexactness of sense perception. In 1921 Carnap wrote about this contrast between sense perception and physical reality with purity and simplicity:

> The total separateness of these two areas cannot be emphasised strongly enough. The first contains the contents of sensation: colours, sounds, smells, pressures, sensations of warmth, etc. — none of which is even mentioned in theoretical physics ... The epistemological question of the relation between the two areas is not at issue here. Whether (along phenomenological-realistic lines) one calls the contents of the first (e.g. the colour blue) 'mere appearances' and those of the second (e.g. the corresponding electromagnetic waves) 'reality' – or vice versa (along positivist lines) one designates those of the first 'the real given' and those of the second 'mere conceptual complexes' – about that question, physics doesn't need to be concerned ... physics expresses itself naturally with the help of purely formal correspondence-relations and leaves such interpretations to a non-physical investigation.[29] [ASP (1921c), 10–12]

It has been suggested by some scholars (Carus and Stein *inter alia*) that even when the two domains of cognitively meaningful statements and the class of expressions of *Lebensgefühl* were regarded by Carnap as mutually exclusive, he still left room for essentially *practical external principles*. [*ibid.*] These are connected with what Hermann Helmholtz called "metaphysical hypotheses", which are applied to provisionally complete the picture provided by science in order to cover those aspects of reality that are not yet accessible to it. Moreover, these hypotheses are also necessary as practical guides in life:

> These hypotheses are even more essential for practical action [*das Handeln*] [than for science itself], because one can't always wait around for a secure

[29]Cited in [Carus (2007), 124]

> scientific decision to be reached, but must make a decision, whether according to probability or according to aesthetic or moral feeling. In this sense as well, there can be no more objection to metaphysical hypotheses. [Helmholtz (1878a) [PT, 360]]

The expression of this attitude of Helmholtz was not restricted to a single lecture, but was an overarching theme of his philosophy. He was especially interested in the relation between the physical sciences and the moral sciences, and pondered seriously about the conceptual underpinnings of both. In such enquiries he critically evaluated the intellectual tradition stemming from Hegel, but rather than rejecting the views brought forth within it, attempted to discern in them what would be an adequate basis for a reconciliation between the physical and moral sciences which had dishearteningly deteriorated during the nineteenth century. Nevertheless, one cannot avoid thinking that an insurmountable cleft was still left open between the two domains. As Helmholtz remarked in his popular lectures on the relation of natural science to general science:

> In all branches of those studies, in theology, politics, jurisprudence, æsthetics, philology, there started up enthusiastic Hegelians, who tried to reform their several departments in accordance with the doctrine of their master, and, by the royal road of speculation, to reach at once the promised land and gather in the harvest, which had hitherto only been approached by long and laborious study. And so, for some time, a hard and fast line was drawn between the moral and the physical sciences; in fact, the very name of science was often denied to the latter.
>
> The feud did not long subsist in its original intensity. The physical sciences proved conspicuously, by a brilliant series of discoveries and practical applications, that they contained a healthy germ of extraordinary fertility; it was impossible any longer to withhold from them recognition and respect. And even in other departments of science, conscientious investigators of facts soon protested against the over-bold flights of speculation. Still, it cannot be overlooked that the philosophy of Hegel and Schelling did exercise a beneficial influence; since their time the attention of investigators in the moral sciences had been constantly and more keenly directed to the scope of those sciences, and to their intellectual contents, and therefore the great amount of labour bestowed on those systems has not been entirely thrown away.
>
> We see, then, that in proportion as the experimental investigation of facts has recovered its importance in the moral sciences, the opposition between them and the physical sciences has become less and less marked. Yet we must not forget that, though this opposition was brought out in an unnecessarily exaggerated form by the Hegelian philosophy, it has its foundation in the nature of things, and must, sooner or later, make itself felt. It depends partly on the nature of the intellectual processes the two groups of sciences involve, partly, as their very names imply, on the subjects of which they treat. [Helmholtz (1895) [PT, 7–8]]

In addition to the influence of Helmholtz and Poincaré who played a prominent role in the formation of Carnap's thought on the problems pertaining to philosophy of space, Carnap was strongly influenced in the early 1920s by an original neo-Kantian, Hans Vaihinger (1852–1933), whose *The Philosophy of As If (Die Philosophie des Als-Ob)* made a lasting impression on him. Vaihinger was known as a prominent Kant scholar, having written a massive two-volume commentary on the *Critique of Pure Reason* as well as having founded the prestigious journal *Kant-Studien* in 1896. The latter has remained a prominent forum for Kant scholarship to this date. Vaihinger, regarding himself as a Kantian, and a disciple of Lange (1828–1875), was persuaded in his thinking by quite diverse ideas the most prominent of which had connotations of both Darwinism and pragmatism. Indeed, he maintained that human intellectual life was subservient to practical ends, and therefore to (human) passions. Being an enthusiastic Darwinian, he saw the culturally evolved tools of language and thought as essentially survival mechanisms ("purposefully operating organic functions") developed in natural evolution. The products of these "organic functions", viz. theories and concepts, were to be understood within the overall system of economically regulated human organism, essentially having practical significance for the agency of the entire organism. Hence, the traditional picture of concepts and theories as vehicles of accurate representation of something external to the organism was insufficient, comprising a secondary task only, subsumed under the primary task of survival and attaining practical ends. This overall philosophical vision is put by Vaihinger with rectitude and delicacy in the beginning of the *Die Philosophie des Als-Ob*:

> Das wissenschaftliche Denken ist eine Funktion der Psyche. Unter "Psyche" verstehen wir zunächst nicht eine Substanz, sondern die organische Gesamtheit aller sogen. "seelischen" Aktionen und Reaktionen; diese fallen niemals unter die äussere Beobachtung, sondern müssen teils aus physischen Merkmalen erschlossen, teils mit dem sogen. inneren Sinne beobachtet werden. Die psychischen Aktionen und Reaktionen sind, wie alles uns bekannte Geschehen , notwendige Vorgänge, d. h. sie folgen mit zwingenden Regelmässigkeit aus ihren Bedingungen und Ursachen; will man die psychischen Vorgänge mit einem Gebiet des äusseren Geschehens vergleichen, so eignen sich dazu weniger die physikalischen und im engeren Sinne mechanischen Vorgänge, als die Funktionen des Organismus. [...] Diese Behauptung findet ihre Begründung in dem Umstand, daß wie bei den organischen Funktionen der leiblichen Sphäre, so auch bei den psychischen Funktionen sogen. empirische Zweckmässigkeit beobachtet wird. Diese Zweckmässigkeit äussert sich hier wie dort in einer geschmeidigen Anpassung an die Umstände und an die Umgebung; in einer, die Erhaltung des physischen oder psychischen Organismus anstrebenden und erreichenden Reaktion auf äussere Anstöße und Einwirkungen; in der Aneignung und Aufnahme oder Aubstoßung neuer Elemente. In der Psyche findet nicht bloß ein mechanisches Spiel von Vorstellungen statt, sondern die Vorstellungsbewegung erfüllt in ihrer stetigen Abänderung in hohem Grade die Anforderungen der Zweckmässigkeit. [...] Sämtliche psychischen Prozesse sind in

> dem angegebenen Sinne Zweckmässig; vor allem aber partizipieren an dieser Zweckmässigkeit die sogen. theoretischen Apperzeptionsprozesse. Das wissenschaftliche Denken besteht in solchen apperzeptiven Prozessen, es ist daher unter dem Gesichtspunkt einer organischen Funktion zu betrachten. [Vaihinger (1927), 1–2]

Vaihinger was seen by many as an intellectual ally of the American pragmatists.[30] Ironically enough, he had a low opinion of William James. Vaihinger found James' theory of truth a bitter pill to swallow as it did not distinguish between truth and belief, a distinction that was essential to Vaihinger's own philosophy. Beliefs, according to Vaihinger, are useful fictions, akin to Kant's ideals. We believe them in order to attain practical ends (or minimally, to give us at least some prescription for action instead of sustaining all beliefs and refraining from action). We believe them while realizing that they may well turn out to be false. The fictions are necessary tools in acquiring knowledge about the world: without fictions we are locked in a chaotic and utterly formless subjectivity of the present moment. [Carus (2007), 126] The ultimate measure of fictions is their usefulness in the conduct of life. The ingredient of skepticism in Vaihinger's thought winds about the idea that even if some of our useful fictions in science, for example, were actually true, we could not have any way of knowing that.

In the following sections I intend to show the central significance of geometry and logico-geometrical concept formation to Carnap's early constitutional program of philosophy. This early stage of the constitutional program is seen to be of utmost importance to the later attempts of Carnap to refine the notion of "rational reconstruction". The apogee of this development is the method of explication introduced in 1950 in the *Logical Foundation of Probability*. Indeed, the geometrical origins of the method become apparent, when one notices the quite unusual sense of geometry that Carnap has recourse to in his early studies. I intend to shed light on this particular issue. I will also briefly consider what kind of ontological consequences this approach has for Carnap. The examples are mainly taken from physics.

4.3 Foundational Studies in Geometry

Carnap's philosophical career did not begin with a flying start. In the midst of his studies at the universities of Jena and Freiburg the First World War broke out and Carnap was recruited to the Prussian army. He served as a communications engineer at the front and was wounded in action. Afterwards he was decorated. It is quite surprising that even the ordeal of experiencing the horrors of war first-hand did not hamper Carnap

[30] Klaus Ceynowa has argued that there is a mutual ancestor to the philosophies of American pragmatism and Vaihinger: the psychological theories of Alexander Bain. As is well known, they were directly adopted by both James and Peirce, but their transmission to Germany was more indirect, having been conveyed there by the neurophysiologist Adolf Horwicz. Vaihinger acknowledged him as an important influence. [Ceynowa (1993)]

from enthusiastically absorbing himself in intellectual pursuits. The following statement from the Intellectual Autobiography [Carnap (1963), 3–84] reveals in particular his early acquaintance with Einstein's theory of relativity.

> Even during the war, my scientific and philosophical interests were not entirely neglected. During a quiet period at th Western Front in 1917 I read many books in various fields, e.g., about the world situation and the great questions of politics, problems of *Weltanschauung*, poetry, but also science and philosophy. At that time I became acquainted with Einstein's theory of relativity, and was strongly impressed and enthusiastic about the magnificent simplicity and great explanatory power of the basic principles. Later, in Berlin, I studied the theory of relativity more thoroughly and was especially interested in the methodological problems connected with it. I also wrote circular letters about the theory to a few friends; I included an article or small book by Einstein or others and added detailed explanations with diagrams. Thus I tried to share my great intellectual enjoyment of the theory with friends. [Carnap (1963), 10]

The early interest in the theory of relativity instigated Carnap to study geometry in depth for, as he realized, the methodological questions pertaining to the former had very much to do with the latter. Furthermore, the philosophical import of these issues was considerable. Carnap's occupation with the philosophical issues pertaining to geometry was quite typical at the time and reflected the general orientation of intellectual activity in the academia. Indeed, investigations into the conceptual foundations of geometry were actively pursued by mathematicians, physicists and philosophers alike. Especially notable among these investigations were the ones by Klein and Poincaré. In addition to resulting in masterful presentations of the requisite concepts and methods in making the mathematical study of geometry both more rigorous and general than it had been before, they also exhibited awareness of and impeccable taste in regard to the philosophical issues that related to them. Poincaré in particular maintained the intimate connection between theoretical geometry and questions about human experience that was the hallmark of the Kantian notion of the "pure form of our sensible intuition". But equally enlightening are the examples of the remarkable incentive that geometry formed in the thought of several philosophers. A few specific examples should suffice: (i) It is well known that Bertrand Russell's first book was concerned with these questions. Indeed, this book — in fact, his dissertation in philosophy — was titled *Foundations of Geometry* and was published in 1897. Shortly after that he published another book, *The Foundations of Mathematics*, in which the issues pertaining to geometry figured prominently. (ii) Ernst Cassirer is known to have valued Klein's *Erlangen Programm*[31] highly, for he envisaged it as a guideline for

[31] The content of Klein's program can be summarized in the following, intuitive, non-rigorous way: Klein proposed to view geometry as "the study of the properties of a space which are invariant under a given group of transformations". To study geometry, he said, one needed not only objects (triangles, circles, icosahedra, or much wilder things [fractals, or figures of fractional dimension, we could say today]), but also movements. In the classical Euclidean regime which had been around for over two millennia, these movements had always been rigid motions: pick up a figure and place an identical copy down in a new place.

his critical idealism. Indeed, in *Substanzbegriff und Funktionsbegriff* Cassirer devotes an entire chapter to the question of concept formation in geometry. The reason for this is philosophically transparent: Cassirer saw in geometry a paradigmatic model of concept formation in science in general. This was not merely a refractory view deriving from the neo-Kantian circles where geometry had a time-honoured place among the sciences, but was symptomatic of the genuinely current tensions and pressures that affected the conceptual foundations of geometry. (iii) Finally, Rudolf Carnap's contributions in this field were highly original. Carnap's dissertation, *Der Raum, Ein Beitrag zur Wissenschaftslehre*, is a remarkable little book, acute in its conceptual approach and up-to-date in the problems it tackles. Strangely, it has been all but neglected among the scholars interested in the philosophy of geometry and philosophy of science in general, and Carnap-scholars in particular. Only a few reasonably sustained analyses of the book exist. Among these the most important ones are Thomas Mormann's articles about *Der Raum* and its position in the foundational discourse about geometry in the first decades of twentieth century, and Guillermo E. Rosado Haddock's provocatively titled *The Young Carnap's Unknown Master — Husserl's Influence on* DER RAUM *and* DER LOGISCHE AUFBAU DER WELT. I will adduce the main ideas of this important early work after laying some conceptual and historical background.

4.3.1 A sketch of the history of non-euclidean geometries

I will briefly adduce the discoveries in mathematics leading to non-Euclidean geometry to set the stage for the analysis of the philosophical issues that lie at the heart of these mathematical questions. The issues involved comprise a fascinating web of ideas and problems. First of all there is the mathematical side to the story, consisting of the history of the discovery and first developments of non-Euclidean geometry, where Gauß plays the part of the originator with his initial researches motivated by his interest in the foundations of geometry. This line of development reaches its culmination in Riemann's memoir of 1854. Another wave of research gains momentum after that with Beltrami traveling upon its crest. His *Saggio* of 1868 marks the beginning of a new approach to these problems.[32] Finally, intertwined with all these technical developments are the philosophical questions pertaining to the nature of space, its ontological status, and our knowledge about its form. Let me begin with the discovery of *hyperbolic geometry*. Christian Houzel reports that this variant of geometry that ultimately became known as the geometry of Lobachevsky, "was discovered several times independently during a period that extends from 1816 to 1829". [Houzel (1992), 5] But similar efforts were made earlier, albeit from a differently motivated basis. For instance, in the 18th century the most conspicuous results in this vein could be found in the works of Saccheri and Lambert. Their aim was first and foremost to prove the fifth postulate of Euclid, and in that way to vindicate the Euclidean theory of parallels. The method they used was tantamount to *reductio*

Klein's radical idea was that other movements, which might stretch or twist the objects quite drastically, could be thought of as geometrical movements too. [Mumford, Series & Drive 2002, 1–2]

[32] Incidentally, Beltrami's work proved crucial for Einstein's discovery of the correct form of the gravitational field equations.

ad absurdum. This meant that a correlative to the fifth postulate was posited by negating the latter. This postulate could naturally take various forms. Applying similar non-Euclidean hypotheses in their geometrical derivations, Saccheri and Lambert drew very different morals from their results. Saccheri found the implication of the result that non-Euclidean parallels must be *asymptotic*, namely their having a common perpendicular at their point of intersection at infinity, *repugnans naturae lineae rectae*, that is, "repugnant to the nature of the straight line". [*ibid.*] Lambert did not concur. His opinion was rather that the non-Euclidean hypotheses were difficult to refute, and hence one should be more careful in judging the truth of the non-Euclidean geometries. However, as his motivation was to vindicate the Euclidean viewpoint epitomized in the fifth postulate, he refrained from publishing his results that were collected in his book *Theorie der Parallellinien*.

About a century later Riemann presents his fundamental new insights into the nature of geometry. A philosophical motivation for Riemann's work derives from J. F. Herbart's theorizing about so-called *serial forms* (*Reihenformen*). This motivation led to the crucial invention of the notion of an n-*ply extended manifold*. In this Riemann relied on the similar notions of Grassmann and Schläfli, which he elaborated heavily . His approach was considerably broader and more general, and made it possible to expand on the variety of measure-relations that could be of use in both pure and applied geometry. The important research questions for Riemann were, then, the number and type of distinct "measure-relations that such a manifold is capable" and the possibility "to express geometrically the calculated results." [Riemann (1868) [Ewald (1996), 652–661]][33] Building on the foundations laid down by Gauß in the celebrated memoir *Disquisitiones generales circa superficies curvas*, Riemann could give an outline of his geometric program in the paper *On the hypotheses which lie at the foundation of Geometry* [*Über die Hypothesen, welche der Geometrie zu Grunde liegen*]:

> Position-fixing being reduced to quantity-fixings, and the position of a point in the n-ply extended manifold being consequently expressed by means of n variables $x_1, x_2, x_3, \ldots, x_n$, the determination of a line comes to the giving of these quantities as functions of one variable. The problem consists then in establishing a mathematical expression for the length of a line, and to this end we must consider the quantities x as expressible in terms of certain units. I shall treat this problem only under certain restrictions, and I shall confine myself in the first place to lines in which the ratios of the increments dx of the respective variables vary continuously. We may then consider these lines broken up into elements, within which the ratios of the quantities dx may be regarded as constant; and the problem is then reduced to establishing for each point a general expression for the linear element ds starting from that point, an expression which will thus contain the quantities x and the quantities dx. I shall suppose, secondly, that the length of the linear element, to the first order, is unaltered when all the points of this element undergo the same infinitesimal displacement, which implies at the same time that if all the quantities dx are increased in the same ratio, the linear element will vary also in the same

[33]Clifford's translation; revisions by William Ewald.

> ratio. On these suppositions, the linear element may be any homogeneous function of the first degree of the quantities dx, which is unchanged when we change the signs of all the dx, and in which the arbitrary constants are continuous functions of the quantities x. To find he simplest cases, I shall seek first an expression for manifoldness of $n-1$ dimensions which are everywhere equidistant from the origin of the linear element; that is, I shall seek a continuous function of position whose values distinguish them from one another. In going outwards from the origin, this must either increase in all directions or decrease in all directions; I assume that it increases in all directions, and therefore has a minimum at that point. If, then, the first and second differential coefficients of this function are finite, its first differential must vanish, and the second differential cannot become negative; I assume that it is always positive. This differential expression, then, of the second order remains constant when ds remains constant, and increases in the duplicate ratio when the dx, and therefore also ds, increase in the same ratio; it must therefore be ds^2 multiplied by a constant, and consequently ds is the square root of an always positive integral homogeneous function of the second order of the quantities dx, in which the coefficients are continuous functions of the quantities x. For Space, when the position of points is expressed by rectilinear co-ordinates, $ds = \sqrt{\sum(dx)^2}$. [*ibid.*, 655-656]

A few comments are in order here. By "Space" Riemann means three-dimensional Euclidean space. The systematic consideration of different possible cases of dependence between the quantities ds and dx leads Riemann finally to restrict the investigation to "manifolds for which the line element is given by the square root of a differential expression of the second degree", namely

$$ds^2 = \sum_i \sum_j g_{ij} dx_i dx_j,$$

where $g_{ij} = g_{ji}$ are functions of the variables x_i. This requirement is tantamount to a guarantee that the n-dimensional space is locally Euclidean. This characterization of the local geometry of an n-manifold is, as we shall see, an essential ingredient of the general theory of manifolds.

The essential feature of surfaces that the Riemannian formulation of geometry enables one to tackle with is their curvature. Perhaps it is expedient to give an informal description of the procedure of defining the principal curvature of a given surface. Without making the definitions exact, we can give the following outline of the recipe for calculating the curvatures κ_1 and κ_2. Suppose S is a surface in $\mathbb{R}^3$, p is a point in S, and N is a unit normal vector to S at p. Then, to calculate κ_1 and κ_2:

1. Choose a plane Π through p that contains N. The intersection of Π with S is then a plane curve $\gamma \subset \Pi$ passing through p.

2. Compute the signed curvature κ_N of γ at p with respect to the chosen unite normal N.

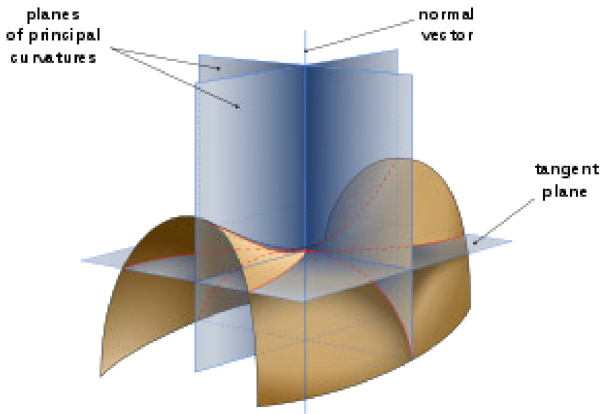


Figure 4.1: Principal curvatures. ©Wikipedia.

3. Repeat this for *all* normal planes Π. The *principal curvatures* of S at p, denoted κ_1 and κ_2, are defined to be the minimum and the maximum signed curvatures so obtained.

4.3.2 The conceptual foundations of geometry

In the wake of the researches by Klein, mathematicians came to appreciate the pluralistic nature of synthetic geometry. Discarding the dichotomy between Euclidean and non-Euclidean geometries and the suitability of the one or the other in describing the structure of our spatial perception as the central question of synthetic geometry, a vast field of inquiry was suddenly opened up. The radical insight was that synthetic geometry itself provided a wealth of approaches to formulate the axiomatic basis of geometry. One of the most distinctive novelties was the introduction of the idea of finite geometries. In these systems one deals with only a finite number of basic geometrical entities, principally points and lines. A famous example of these is Fano's system of geometry, which is a (formally symmetrical) system of only seven lines and seven points. Another crucial insight was the idea of duality which asserts: Assume S to be a geometric system in which points and lines are related to each other in a certain way. Then there exists a geometrical system S^* in which the points of S play the functional role of the lines in S^* and the lines of S correspond to the points of S^*. This means, in other words, that points and lines are always defined interdependently: knowledge about the points of a geometric system entails knowledge about its lines.

These considerations change radically the status of geometry in relation to the Kantian question of the transcendental conditions for the possibility of spatial experience and perception. But a valid kernel of the Kantian conception remains untouched: the very principle of imposing such transcendental conditions on our experience. Only now the centrality of Euclidean geometry is questioned. Thomas Mormann sheds light on the core ideas of the newly conceived foundations of synthetic geometry. [Mormann (2003), 47] He emphasizes that the *Leitmotiv* of synthetic geometry is the concept of order. Incidentally, also Carnap conceived geometry as a general theory of *Ordnungsgefüge*. Carnap intended this term in a semi-technical sense, implying by it something akin to "relational structure" or "structured set".[34] By stipulating order on a domain (*Ordnungssetzungen*), one conceives it as an *Ordnungsgefüge*. In general, the point of geometry is to study the variety of different geometries. The two characteristics of this approach to geometry have been summarized by Mormann as follows:

1. Space in the sense of synthetic geometry is a general term which comprises many different spatial structures. Geometry has to study all of them without blinders to single one as the "true" one. In this sense geometry is abstract, not in that it is remote from applications.

2. Synthetic Geometry is relational. Its objects are determined by a net of implicit relational definitions. The ontological status of any geometric object is determined by its relational position within a certain relational system. [*ibid.*, 48]

Another line of research was prompted by David Hilbert whose penetrating researches into the foundations of geometry, especially in his *Grundlagen der Geometrie*, paved the way for purely formal treatments of synthetic geometry. In addition, Hilbert was also influential in instigating new insights into the nature of plane curves, anticipating the theory of fractals.

4.3.3 Carnap's analysis of space: *Der Raum*

In addition to the logical considerations mentioned above, the theory of relativity played a prominent role in shaping the discussion about the foundations of geometry. Carnap's dissertation attempted to combine both of these approaches and, to a certain degree, succeeded in it. The fundamental insight of his dissertation was the introduction of a division of different kinds of spaces in order to demarcate different discourses from one another. Bruno Bauch's course on Kant's *Critique of Pure Reason* where the conception that the geometrical structure of space is dictated by the form of our intuition was adduced, made Carnap to pursue firmly the philosophical task of clarifying the underpinnings of geometrical concepts and their varied use in mathematics, physics and philosophy. Carnap maintained that while the same term "space" was used in all these disciplines, each of them accommodated entirely different subjects under their purview:

> [. . .] I distinguished three meanings of this term, namely, formal space, intuitive space, and physical space. Formal space is an abstract system, constructed in mathematics, and more precisely in the logic of relations; therefore our knowledge of formal space is of a logical nature[35] Knowledge of in-

[34]As should be clear from this characterization, it is not a *terminus technicus* in geometry.

[35]In 1919 Carnap had studied the magnum opus of Russell and Whitehead, the *Principia*, and "was strongly impressed by the development of the theory of relations in this work" [Carnap (1963), 11].

> tuitive space I regarded at that time, under the influence of Kant and the neo-Kantians, especially Natorp and Cassirer, as based on "pure intuition" and independent of contingent experience.[36] But, in contrast to Kant, I limited the features of intuitive space grasped by pure intuition to certain topological properties; the metrical structure (in Kant's view, the Euclidean structure) and the three-dimensionality I regarded not as purely intuitive, but rather as empirical. Knowledge of physical space I already as entirely empirical, in agreement with empiricists like Helmholtz and Schlick. In particular, I discussed the role of non-Euclidean geometry in Einstein's theory. [Carnap (1963), 12]

Carnap saw clearly the importance of the concept of a topological space for the rigorous development of the geometrical foundations of the general theory of relativity in particular, and for the foundations of geometry in general. It was therefore clear at the outset that the classification of spaces should include topological spaces as one of the main categories. But there is another, conceptually more compelling reason. It is simply that the notion of a topological space enables us to proceed in abstraction a little further from the concept of a metric space that still contains some "extraneous" information. This kind of an abstraction would make it possible, Carnap thought, to retain the Kantian idea of the transcendental condition of our spatial experience in a more relaxed form. Reverting to the mathematical aspects of the idea, it is clear that a homeomorphism between two metric spaces need not preserve distances (a paradigmatic example is a homeomorphism between two spheres with different radii). Mathematicians therefore introduce the concept of a "space" without distances in which continuous functions still make sense. Thus we end up with the following definition of a topological space:

A *topology* on a set X is a collection $\mathcal{T}$ of subsets of X, called *open sets*, satisfying the following properties

(i) X and $\emptyset$ are elements of $\mathcal{T}$.

(ii) $\mathcal{T}$ is closed under finite intersections: If $U_1, \ldots, U_n \in \mathcal{T}$, then their intersection $U_1 \cap \cdots \cap U_n$ is in $\mathcal{T}$.

(iii) $\mathcal{T}$ is closed under arbitrary unions: If $\{U_\alpha\}_{\alpha \in \mathcal{A}}$ is any (finite or infinite) collection of elements of $\mathcal{T}$, then their union $\cup_{\alpha \in \mathcal{A}} U_\alpha$.

A pair $(X, \mathcal{T})$ consisting of a set X and a topology $\mathcal{T}$ on X is called a *topological space*.

The concept of a topological manifold is modeled on the concept of topological space. More exactly, a topological manifold is a second countable Hausdorff space locally homeomorphic to a Euclidean space $\mathbb{E}^n$. By imposing a definite metrical structure (a *Riemannian metric*) on the topological manifold $\mathcal{M}$ and demanding in addition that it be *smooth*,

[36] Curiously enough, Carnap does not here mention Husserl, whose impact on the overall view of *Der Raum* and especially on the parts concerning the intuitive space was conspicuous. The reasons for this "selective amnesia", as [Haddock (2008)] calls it, are not entirely clear.

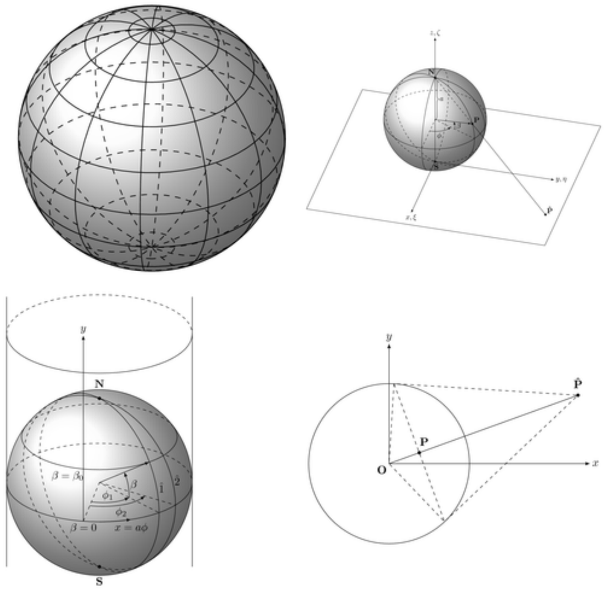
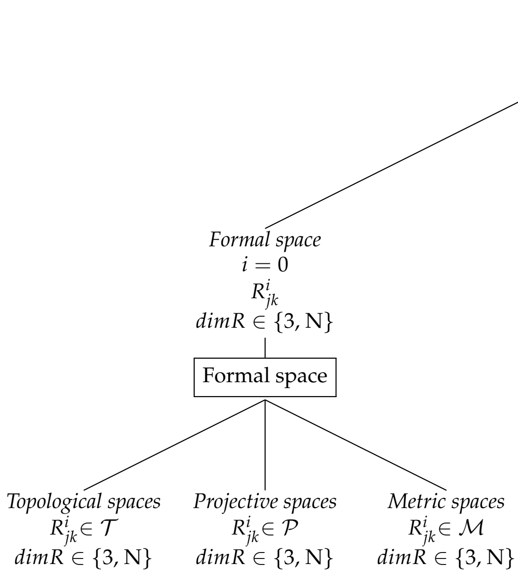


Figure 4.2: Map-projections. ©Tomasz M. Trzeciak. Source: `LaTeX-Community.org`.

i.e. that the function defined on $\mathcal{M}$ are $\mathcal{C}^{\infty}$, that is to say, infinitely differentiable, one ends up with a type of manifold that forms the basic tool in the general theory of relativity (*Riemannian* or *pseudo-Riemannian manifold*).

Carnap made a subdivision in his classification of spaces consisting of a mathematical division into *topological*, *projective* and *metrical* spaces. The meaning of the concept of topological space we have adduced above. The two remaining types of space can be defined in various ways. (In the following I will employ just one particular definition of them, and refrain from a detailed explanation of these concepts.) Carnap actually confused the notions of projective and affine space,[37] but this is not fatal for his treatment of the general problem. The basic idea is that the notion of projective space was to play an intermediate role between the topological space, where no metrical structure could be specified and the metrical space, where such structure could be given by ascribing a distance function, a *metric* to the space. In an intuitive language, *projective geometry is concerned with those properties of figures that are left unchanged when the figures are subjected to a projective transformation*. Illustrations of projective transformations in Euclidean space are given in the above figure.

[37]Mormann in a personal communication to Haddock [Haddock (2008), 6n].



This amounts to 18 different kinds of space in all.

Let us now specify in a little more detail the ideas behind this classification. In the first place, Carnap means by formal space simply a relational structure (*Ordnungsgefüge*) whose members lack of any determination (as elements of the space), and which are characterized only by the most abstract relational properties of geometry. In practice such a formal specification means that from certain connections of a determined kind one can draw conclusions about connections of another determined kind within the same region. In short, this class includes the purely mathematical aspects of spatial description. The nature of the intuitive space is, however, something at once more concrete:

> By intuitive space, on the other hand, is understood the structure of relations between "spatial" figures in the usual sense [...] whose determined particularity we apprehend by means of perception or mere representation. One is still not concerned there with spatial facts present in empirical reality, but only with "essence" of those figures, which can be recognized in any representative of the species. [Carnap (1922), 5–6][38]

The physical space, the elements of which comprise the material bodies, particles and fields described by physics, and the existence of which is an empirical fact, presupposes, in a purely logico-epistemological sense, the intuitive space, which, on the other hand, presupposes the formal space. In Carnap's words, "Cognitions of physical space presuppose the cognition of intuitive space, and the latter [...] finds the pure form of its

[38]Translation in [Haddock (2008), 5]. The German original reads: "Unter Anschauungsraum dagegen wird das Gefüge der Beziehungen den im üblichen Sinne "räumlichen" Gebilden verstanden, also des Linien, Flächen- und Raumstücken, deren bestimmte Eigenheit wir bei Gelegenheit sinnlicher Wahrnehmung oder auch bloßer Vorstellung erfassen. Dabei handelt es sich aber noch nicht um die der Erfahrungswirklichkeit vorliegenden räumlichen Tatsachen, sondern nur um das "Wesen" jener Gebilde selbst, das an irgendwelchen Artvertretern erkannt werden kann."

structure prefigured in formal space." [*ibid.*, 7] More specifically, the theories of formal and intuitive space stand in the relation of specification, and the theories of intuitive and physical space stand in the relation of subordination. This hierarchy is, then, quite clearly seen to correspond to the Husserlian one specified in the *Logische Untersuchungen*, which comprises the relations between formal ontology, regional ontology and factual science. [Husserl (1970) [1900/1901], 303ff.] The alleged philosophical import of Carnap's classification, as he went on arguing, was that it enabled, together with a modification of Kant's original account, a dissolution of the traditional dispute over the Kantian doctrine of the synthetic a priori basis of cognition of space in pure spatial intuition:

> The old controversies between mathematicians, who disputed Kant's assertion, and philosophers, who defended it, were thus obviously unable to reach any result, because the two sides were not talking about the same object. The former had partly formal space in mind (e.g. Couturat) and partly physical space (Riemann, Helmholtz, Poincaré), the latter intuitive space. So both Parties were correct and could have been easily reconciled if clarity had prevailed concerning the three different meanings of space. [*ibid.*, 47]

Carnap provides a detailed presentation of the different sorts of space in the first three chapters of *Der Raum*, whereas the relationship between these different spaces is specified in the fourth chapter. In the fifth and last chapter an overall evaluation of the three conceptions of space is provided. In particular, the formal space (*formaler Raum*) is singled out as an individual chapter in the development of the logical theory of relations as first systematically specified in the works of Russell and Whitehead, especially in Russell's *Principles of Mathematics* and in Russell's and Whitehead's monumental *Principia Mathematica*. Subsuming the theory of formal space in the general theory of relations entails that along the properties of numbers (as specified in number theory), the properties of formal space are, not only independent of experience, but ultimately derivable from logic. Frege's logicism, the influence of which on Carnap was briefly assessed in section **4.1.2**, did not ruminate into the domain of geometry, as his conception of geometry was very much in line with the Kantian one, constituting thus a peculiar tension between the analyticity of arithmetic and the synthetic nature of geometry.

As to the specific content of the notion of intuitive space, Carnap's presentation was heavily influenced by Husserl, as we noted above. This is conspicuous in the remarks in the fifth chapter of *Der Raum*:

> Here [in intuitive space] we have distinguished between the principles in the strict sense and the requirements. Those build the results of a determined sort of "contemplations of essences" [*Wesenerschauung*] (in Husserl's sense) and as such, like all knowledge from this source, do not need the accumulation of empirical facts, [and] as such are not to be referred to as empirical knowledge, but also are not independent of every experience, since they are obtained from any representative of the kind of objects concerned. Requirements, on the other hand, are not knowledge but stipulations that have to

be made in order to obtain a total structure "space" from such knowledge, which in virtue of their own nature [*ihrem Wesen nach*] seem limited to an incomplete region. For these extensions to a complete structure different possibilities were indicated. Topological space presents [*darstellt*] what is common to all [of those possibilities] and on this ground ought to be seen as the form of the spatial apprehensible in the contemplation of essences. Intuitive metric spaces, on the contrary, also depend on the choice of the stipulations; and as such lack the property of the unlimited validity, which possess [both] intuitive topological space and all knowledge originating in this source. [Carnap (1922), 62–63][39]

Thus, the principles of intuitive space, knowable by means of Husserlian intuition of essences [*Wesen*], completely determine intuitive topological space. The primacy of the purely topological properties of space is here also clearly visible. As to physical space, Carnap contends that we come to know its properties empirically, through the factual content of our experience. We generalize upon the most conspicuous features of these contents by induction, and these features, once again, uniquely determine only the physical topological space. The systematic emphasis on the topological properties of space at every level of the classification — formal, intuitive, physical — is a hallmark of Carnap's geometrical conventionalism. The idea originates with Poincaré, whose specific results concerning the 'topological equivalence' of the Euclidean space and a particular hyperbolic representation of it very much influenced Carnap's thinking in this vein. In a nutshell, Poincaré's construction establishes an equivalence between the three-dimensional Euclidean space and a ball centered at origin, the radius of which is R, and the metric inside which is stipulated by a specific rule of dilatation:

> Suppose [. . .] a world enclosed in a large sphere and subject to the following laws. The temperature is not uniform; it is greatest at the centre and gradually decreases as we move towards the circumference of the sphere, where it is absolute zero. The law of this temperature is as follows: if R be the radius of the sphere and r the distance of the point considered from the centre, the absolute temperature will be $R^2 - r^2$. Further, I will assume that in this world all bodies have the same coefficient of dilatation, so that the linear dilatation of

[39] "Hier haben wir unterschieden zwischen den Grundsätzen im engeren Sinne und den Forderungen. Jede bilden den Befund einer bestimmten Art der "Wesenerschauung" (im Husserlschen Sinne) und sind daher wie alle Erkenntnisse dieser Quelle nicht auf Häufung von Erfahrungstatsachen angewiesen, daher nicht als Erfahrungserkenntnisse zu bezeichnen, aber auch nicht unabhängig von jeder Erfahrung, insofern als sie an irgendwelchen Vertretern der betrefflichen Art von Gegenständen gewonnen werden. Die Forderungen dagegen sind nicht Erkenntnisse, sondern Festsetzungen, die getroffen werden, um ein geschlossenes Gesamtgefüge "Raum" aus jenen Erkenntnissen zu gewinnen, die ihrem Wesen nach auf ein nicht vollständiges Gebiet beschränkt erscheinen. Für diese Erweiterungen zum vollständigen Gefüge zeigten sich verschiedene Möglichkeiten. Der topologische Raum stellt das ihnen allen gemeinsame dar und ist deshalb als Form des in der Wesenerschauung des Räumlichen faßbar anzusehen. Die metrischen Anschauungsräumen dagegen sind auch noch von der Wahl jener Festsetzungen abhängig; daher fehlt ihnen die dem topologischen Anscahuungsraum wie allen dieser Quelle enstammenden Erkenntnissen zukommende Eigenschaft der unbedingten Gültigkeit."

> any body transported from one point to another of different temperature is instantaneously in thermal equilibrium with its new environment.[40] [Poincaré (1952) [1902], 65]

The construction of Poincaré had a considerable attraction for Carnap, who attempted to extend the basic idea of that construction by showing that not only was it possible to endow the Euclidean space $\mathbb{E}^3$ with a Riemannian structure of constant *negative* curvature (as was the case with Poincaré's hyperbolic representation), but also with a structure of *positive* curvature (i.e. an elliptic geometry). This procedure got its thrust from the idea that only the topological features of an *Ordnungsgefüge* were matters of fact, whereas the metrical properties of space were merely stipulations, i.e., matters of convention. This, of course, was a leading idea behind the representations of the philosophy of geometry by other logical empiricists, too, and one that would reign the discussions about the foundations of geometry all the way up to the latter half of the twentieth century. Carnap summarized his conception in the following way:

> The transfomation of a statement of matter of fact from one metrical space-form into another — e.g., from the Euclidean one into one of the non-Euclidean — has been aptly compared to the translation of a proposition from one language into another. Now, just as the genuine sense of the proposition is not its presentation in one of these linguistic forms — for then its presentation in the other languages would have to appear as derivative and less original — but is merely that in the proposition which remains unaltered in translation; so too the sense of the statement of matter of fact is not one of its metrical presentations, but that which is common to all of them (the "invariants of topological transformations") — and that is precisely its presentation in merely topological form. [Carnap (1922), 65]

This, then, was the professed relevance of Carnap's work for the philosophical and scientific discourse that centered on the foundational problems of geometry. Although many of his suggestions pertaining to geometry have been contested and some even refuted (such is the case with his thesis that the Kantian conception of *synthetic a priori* can be retained in the way of referring to fundamental topological facts [*topologischer Tatbestand*] underlying any spatial configuration),[41] the lasting influence of his exposition resides in

[40]More specifically, the metric of the Poincaré model is given by:

$$ds^2 = 4\frac{\sum_i dx_i^2}{(1-\sum_i x_i^2)^2}.$$

[41]For example, it is impossible, if it is assumed that the Riemannian metric is complete, to endow the Euclidean space with a constant positive curvature. More generally, the relationship between the topological and metrical structure of a space is far more intricate than the straightforward optimism of Carnap suggests. Indeed, the mathematical progress made in the latter half of twentieth century has enabled us to delve in much more detail to these questions, and the general contention is that all-encompassing generalizations about these relations are in principle impossible to achieve. A convincing example of this is the notori-

the variety of conventionalism that is introduced in that work. This conventionalism, interestingly prefiguring his later ideas of linguistic pluralism, marks the first appearance of one of the main motifs of Carnap's philosophical work.

4.3.4 Reichenbach's *Relativitätstheorie und Erkenntnis A priori*

In his seminal book *Relativitätstheorie und Erkenntnis A priori* about the philosophical significance of Einstein's relativity theory Hans Reichenbach presented an incisive critique of Kantian views of space and time. Although educated under the influence of neo-Kantianism and firmly embedded in its habits of thought — just as Carnap was — Reichenbach came to play a significant role in liberating serious philosophical work pertaining to science and its applications from the restrictive and partly outdated conceptions of Kantian philosophy that were still largely dictating the modes of professional philosophizing in Germany at the time. He was ideally suited to this task for in addition to his impeccable mastery of traditional Kantian philosophy he attended Einstein's lectures on relativity theory at the University of Berlin. Furthermore, Reichenbach's philosophical views on relativity theory had taken their shape in a unique environment: he and Einstein used often to travel to their homes from the university by the same streetcar and continue discussions on the themes adduced in the lecture. Later, when Reichenbach had secured a university position for himself in Berlin by the help of Einstein, these commutes were made with Reichanbach's tiny car[42] and lengthy discussions on relativity would ensue until the trip was over.

Reichenbach, more than giving merely a presentation of the inadequacies of the Kantian approach in assessing the philosophical significance of modern scientific results, tapped right into the core of contemporary concerns about the relationship between philosophy and science. Indeed, he asserted to have introduced a new methodology of investigation which he coined *the method of of logical analysis*. As we have seen, this was not a singular phenomenon at the time. The setting for proposing methodological novelties was provided by the concrete theories and problems of the natural sciences and mathematics. It is evident that Reichenbach's thoughts about the status of the notion of *a priori* and the meaning of the notions of space and time had a crucial influence on the development of Carnap's thought pertaining not only to the problems of the epistemology of geometry but to the nature and extent of scientific philosophy in general. As a matter of fact, the two men were in correspondence from 1919 onwards.[43] However, it was not until 1923

ously recalcitrant problem of deciding the Poincaré Conjecture which was only recently proven by Grigori Perelman. Whereas analogous problems for dimensions of four and higher were solved using similar methods, the Poincaré conjecture involving dimension three was tackled with completely differently (using the so-called Ricci flows familiar from differential geometry).

[42] According to an anecdote told by Maria Reichenbach in the preface to the English edition of *Relativitätstheorie und Erkenntnis A priori*, in a midst of one such drive a policeman at the Brandenburg Gate waved them on with "Get going with your baby-buggy". [Reichenbach (1965), xv]

[43] The Reichenbach-Carnap correspondence comprises approximately two hundred letters and it is contained in the Hans Reichenbach Collection. The Hans Reichenbach Collection is part of the Archives of Twentieth Century Philosophy of Science, which also houses the Rudolf Carnap and Frank Ramsey Col-

that they managed to meet in person at a conference in Erlangen that was organized by them and a few others with similar ambitions in pursuing philosophy from a scientific viewpoint with an emphasis on the conceptual tools provided by mathematical logic.

Carnap was also much influenced by Reichenbach's reading of Kant. Reichenbach had offered a *précis* of the Kantian account of "constitution" of the object of knowledge from the distinct contributions of perception and categories:

> According to Kant, the object of knowledge, the phenomenal thing, is not immediately given. Perception does not give the object but only the material [*Stoff*] of which it constructed through an act of judgement. In judgement a subordination [*Einordnung*] into a determinate schema is carried out, according to the choice of scheme a thing or a determinate type of relation develops. Intuition [*Anschauung*] is the form in which perception represents the material of knowledge; accordingly, intuition contains a synthetic moment. However, only the conceptual scheme, the categories, creates the object [*Objekt*]; the object [*Gegenstand*] of science is therefore not a "thing-in-itself" but rather an intuition-based reference structure [*Bezugsgebilde*], constituted through categories. [Reichenbach (1920), 46; English trans. (1965) 48–49]

To give a summary of Reichenbach's critique of the Kantian analysis of space and time, it is expedient to consider the principles that were essential to Kantianism but demonstrably incompatible with the findings of modern physics in the context of the theory of relativity. The principles in question were formulated by Reichenbach as follows:

- the principle of the relativity of uniformly moving coöordinates
- the principle of irreversible causality
- the principle of action by contact
- the principle of the approximate ideal
- the principle of normal induction
- the principle of absolute time

Reichenbach contended that, given the framework of relativistic physics, these were all false. The juxtaposition between the Kantian and Einsteinian viewpoints sheds considerable light on the dramatic shift in the presuppositions of Newtonian and relativistic physics. The transition from the Newtonian (Kantian) picture to the relativistic one was not, however, as abrupt as this presentation of Reichenbach seems to imply. The actual process of transition was much more graded and some of the most 'radical' aspects of relativity theory were anticipated already by Leibniz. (This is demonstrated by John Earman

lections. The Archives of Twentieth Century Philosophy of Science is located in the Special Collections Department of the University of Pittsburgh's Hillman Library.

in his book *World Enough and Space-Time* [Earman (1989)]) The development of the conception of physical spaces is succinctly captured by the following table of the "classical space-times", which provides a clear view of the differences of the historically important conceptions of space and time since Aristotle: [Earman (1989), 36]

Name	Structure	Symmetries	Invariants
1. Machian space-time	absolute simultaneity; $\mathbb{E}^3$ structure of the instantaneous spaces	$x \to x'^{\alpha} = R^{\alpha}_{\beta}(t)x^{\beta} + a^{\alpha}(t)$ $t \to t' = f(t),\ df/dt > 0$	relative particle distances
2. Leibnizian space-time	(1) + time metric	$x \to x'^{\alpha} = R^{\alpha}_{\beta}(t)x^{\beta} + a^{\alpha}(t)$ $t \to t' = t + \text{const}$	relative particle velocities, accelerations, etc.
3. Maxwellian space-time	(2) + standard of rotation	$x \to x' = R^{\alpha}_{\beta}x^{\beta} + a^{\alpha}(t)$ $t \to t' = t + \text{const}$	rotation for an extended body
4. neo-Newtonian space-time	(3) + inertial structure	$x \to x' = R^{\alpha}_{\beta}x^{\beta} + v^{\alpha}(t) + \text{const}$ $t \to t' = t + \text{const}$	acceleration of a particle
5. Full Newtonian space-time	(4) + absolute space	$x \to x' = R^{\alpha}_{\beta}x^{\beta} + \text{const}$ $t \to t' = t + \text{const}$	velocity of a particle
6. Aristotelean space-time	(5) + distinguished spatial origin	$x \to x' = R^{\alpha}_{\beta}x^{\beta}$ $t \to t' = t + \text{const}$	distance from the center of the universe

Reichenbach's acute analysis of the different meanings of the Kantian notion of *a priori* and its relevance for modern science more than compensated such lapses in historical knowledge, however. The essential novelty of Reichenbach's book was the specification of a system of coördination, i.e., a framework for cognitive systematization, in which the results of modern physical theory could be more adequately described than in the Kantian framework. Indeed, he went on to specify the differences between his and Kant's system of knowledge: "Our view differs from that of Kant as follows: whereas in Kant's philosophy only the determination of a *particular concept* is an infinite task, we contend *that even our concepts of the very object of knowledge, that is, of reality and the possibility of its description, can only gradually become more precise*." [*ibid.*, 88] Moreover, the essential role of the theory of relativity in instigating the reformulation of the character of the theory of knowledge was clearly expressed by Reichenbach:

> If the system of coördination is determined by reason in its conceptual relations, but in its ultimate construction by experience, then the totality expresses the nature of reason and by the reality that the concept is intended to formulate. It is therefore not possible, as Kant believed, to single out in the concept of object a component that reason regards as necessary. The idea that the concept of object has its origin in reason can manifest itself only in the fact that this concept contains elements for which *no* selection is prescribed, that is, elements that are independent of the nature of reality. The arbitrariness of these elements shows that they owe their occurrence in the concept of knowledge altogether to reason. *The contribution of reason is not expressed by the fact that the system of coördination contains unchanging elements, but in the fact that arbitrary elements occur in the system.* This interpretation represents an essential modification compared to Kant's conception of the contribution of reason. The theory of relativity has given an adequate presentation of this modification. [*ibid.*, 88–89]

This passage, essentially containing an elucidation of what could be termed *relativized-a-priori*, is important in pointing not only to the fact that contemporary theories in physics were inspiring interesting work in philosophy, but further to the tendency, already visible also in Carnap's work, of an all-encompassing conventionalism. Given that Carnap and Reichenbach were in contact from early on, it is plausible that they have exerted mutual influence on each other's views about the conventional elements in scientific theories. In any case, Carnap's work after *Der Raum* centered on fundamental issues of physics.

4.3.5 Ontological aspects of Carnap's program

Originally Carnap had planned to write his dissertation on the axiomatics of physics. This kind of work was not, however encouraged in the physics department (Max Wien, Carnap's tutor, had not been overtly enthusiastic about such a research project). Carnap had to settle ultimately with the idea of writing his dissertation about the philosophical foundations of geometry. This did not in the end compromise Carnap's possibilities to study in more detail the conceptual aspects of the theory of relativity that had originally instigated his interest in the axiomatics of physics, because apparently, the questions on the foundations of geometry and their relation to physics were ideally suited for an axiomatic treatment.

The groundwork that Carnap had laid in his dissertation proved fruitful for his consecutive efforts to implement a program aiming towards an ideal system of physics. In the article “Über die Aufgabe der Physik” [Carnap (1923)], published shortly after the revised version of *Der Raum*, Carnap had imagined that this system should consist of three distinct compartments: the first one comprised the basic laws of physics represented as a formal axiom system; the second comprised *the phenomenal-physical* dictionary, that is, a set of correspondence rules establishing the relations between observable qualities and physical magnitudes; and the third one comprised the description of the state of the universe for two distinct and arbitrary time points. What would this kind of an ideal system of physics amount to? As Carnap makes clear, the result was a deterministic picture of physical reality:

> [...] From these descriptions, together with the laws contained in the first volume, the state of the world for any other time-point would be deducible (Laplace's form of determinism), and from this result, with the help of the rules of correspondence, the qualities could be derived which are observable at any position in space and time. [Carnap (1963), 15]

This passage is very revealing in making explicit the commitment to determinism.[44] But in 1923 it was a natural option for one mainly interested in analysing the axiomatic structure of established physical theories and the definability of physical concepts under their purview. After all, Werner Heisenberg presented his version of quantum mechanics in 1925, and it was thus far too early for non-experts to fully realize the implications of the quantum revolution that was under way. Furthermore, the importance of the classical (deterministic) picture of physics for the development of the theory of relativity is well known. Einstein stressed on many occasions his deterministic outlook and the theory of relativity certainly was a manifestation of that outlook.[45] As a matter of fact, the special theory of relativity in particular is the singular exemplar of a physical theory where pure instances of determinism can be found.[46] Hence, it is perhaps not a coincidence that Carnap chose to give a glimpse of his ideal system of physics first in the context of special relativity.

Remarks about time

In [Carnap (1923)] Carnap intended to extend the basic ideas of [Carnap (1922)] to comprise problems about time and causality. It was, then, very much the question about assessing the viability of conventionalism in physics that formed the underlying motivation for the discussion in that paper. Here I do not wish to examine the explicit views

[44]A preliminary definition of *causal* determinism can be given along the following lines: The *world* is *governed by* (or is *under the sway of*) determinism if and only if, given a specified *way things are at a time t*, the way things go *thereafter* is *fixed* as a matter of *natural law*.

[45]A hint of the depth of feeling with which Einstein took a stance with respect to determinism is given by his statement reported in the *Spinoza-Festschrift, 1632–1932* where he points with zealous devotion that the application of determinism to human thinking, feeling and action requires “an unusual purity (*Lauterkeit*), greatness of mind (*Seelengrösse*), and modesty.” [*Spinoza-Festschrift, 1632–1932*, ed. Siegfried Hessing (1962)]

[46]See especially John Earman's account of how the relativistic structure of spacetime improves the fortunes of determinism in [Earman (2007)].

of Carnap in [Carnap (1923)] or its sequels [Carnap (1924), (1925)]. Instead I will briefly sketch the modern outlook on this issue, especially with respect to time. This discussion is a sort of preparation for the exposition on the questions of determinism that follows.

Three basic postulates characterize the conception of time that is familiar from modern relativistic physics. (1) In a given system, time (a given *geometric time*[47]) flows in a unique direction in accordance with the second law of thermodynamics;[48] (2) the time flow can be characterized both from an 'objective' and a relativistic point of view; and (3) causal relations have a temporal direction associated with (1).

Since time has a physically determined direction of flow, it is possible — at least in principle — to identify objectively temporally distinguished states of a given system. Naturally, the choice of a particular, restricted set of observables will affect the way how we identify and label those states, but generally such a distinction is possible and unambiguous in the sense that given two (possibly overlapping, but non-identical) sets of observables, $\mathcal{O}_1$ and $\mathcal{O}_2$, with a fixed procedure for their measurement, the temporal order of any two given events, E_1 and E_2 remains invariant: $t_{\mathcal{O}_1}(E_1) < t_{\mathcal{O}_1}(E_2)$ if and only if $t_{\mathcal{O}_2}(E_1) < t_{\mathcal{O}_2}(E_2)$. Hence, given any frame, we can always find a way of assigning time to particular events in such a way that we can order them temporally in an unambiguous way (excluding some anomalous cases in General Relativity). Now, time flow in a given system can be given both an objective and a mind-dependent interpretation. A couple of classic formulations of the mind-dependence thesis are given by Arhur Stanley Eddington, the eminent astrophysicist, and Hermann Weyl, the illustrious mathematician. Let us first relate the version of Eddington:

> Events do not happen; they are just there, and we come across them. "The formality of taking place" is merely the indication that the observer has on his voyage of exploration passed into the absolute future of the event in question. [Eddington (1920), 51]

Weyl restates the essentials of this view as follows:

> The objective world simply *is*, it does not *happen*. Only to the gaze of my consciousness, crawling upward along the life line of my body, does a section

[47]In General Relativity Theory the assignment of time in a given frame is, in general, a complicated process and consists of a choice between distinct geometric times. A (plain, affinely parameterized, or parameterized) time for general relativity is called *geometric* if it satisfies the following conditions. (i) Its domain of definition is closed under the action of $\mathcal{D}(\mathcal{V})$ on the space of solutions. (ii) If g and g' are in the domain of the time and $g' = d^*g$ for some diffeomorphism $d : V \to V$, then the foliation assigned to g' is the image under d^{-1} of the foliation assigned to g (if the time is affinely parameterized, then we require that such a d preserve the time difference between any two instants; if the time is parameterized, then we require that such a d map the instant labeled by t to the instant labeled by t).

[48]Not that the second law of thermodynamics provides an *implicit definition* of time. I take the second law only to indicate the direction of time, or "time's arrow", as it has become to be called. A. S. Eddington introduced the phrase in 1928 in his Gifford lectures. Eddington had in mind only the singular time-asymmetric phenomenon of the increase of entropy, but in 1979 Roger Penrose distinguished as many as seven different 'arrows'. [Savitt (1995), 4]

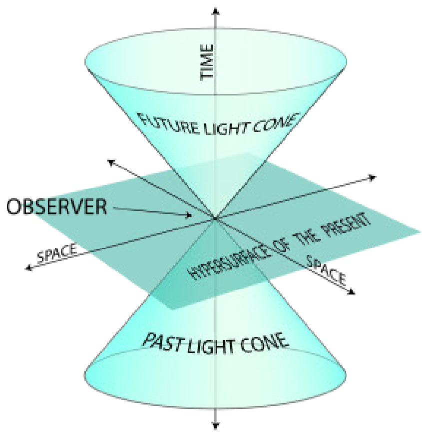


Figure 4.3: Spacetime and the orientation of the light-cone with respect to time axis. ©Wikipedia.

> of this world come to life as a fleeting image in space which continuously changes in time. [Weyl (1949), 116]

Adolf Grünbaum has pondered in this context over the question: what it is that over and above its occurrence at a stipulated, physical clock-time t, characterizes a physical event E as *now* or as belonging to the present? Grünbaum gives his answer along the lines of the mind-dependence thesis: a necessary condition for E to be so characterized is that at t some mind-posessing organism M is conceptually aware of experiencing either E or some other event simultaneous with it in M's reference frame. [Grünbaum (1971), 206–207] This view leads to relativity with respect to the direction of time, but it in no way compromises the general principles of time assignment given above. Indeed, there is no observer-independent label indicating which way is the future direction of flow, but once a choice has been made on the basis of an observed anisotropy (due to irreversible or other time-asymmetric processes), the observer looking at the universe will see a structural difference along the chosen time axis. The selection of the direction of past (and future) is truly a mind-dependent choice. Hence, one can argue that a selection of a time measuring procedure enables one to distinguish between past and future, but that this is only a consequence of a situated decision of a conscious agent; in the physical space-time there does not exist a qualitative difference between past, present and future.

Objective interpretations of time flow intend to dispute the mind-dependence theories by stipulating that time flow is relativistic only in the sense that it manifests itself in a frame-dependent, but not in a mind-dependent way. Without going into the intricate details pertaining to these theories, we can relate their logical implications by giving an outline of their semantics.[49] The most important aspect of the semantics is its temporal definition of truth, which I will give shortly. It is contrasted with the definition of timeless truth that reads:

[49]Details of one such a theory are given by Storrs McCall [McCall (1976) [Tooley (1999), 257–282]]

(A) If a sentence S is true at any time, then either S or a past or future-tensed variant of S is true at any other time.

In the temporal theory of truth we stipulate instead,

(B) If an event-sentence S is true at a certain time t, then either S or a past-tense variant of S is true at all later times. S or a future-tense variant of S may or may not be true at times earlier than t.

As should be clear from the definitions, according to the theory of timeless truth the law of bivalence[50] applies to all event-sentences. In contradistinction, according to the temporal theory, the law of bivalence fails for certain event-sentences given that their assertion times are earlier than their times of reference. To illustrate how strong the truth conditions in Storrs Mc Call's theory [McCall (1976)] of semantics constructed on the basis of (B) are, it is expedient to consider the following account of Mc Call:

> Consider the following example:
>
> **(1)** The earth *collides* with a comet in the year 2000.
>
> If, as is quite possible, the position and velocity of the earth and the comet in 1976 are such that they are on a collision course for the year 2000, and if no outside influences such as rocket motors mounted on the earth can alter this, then the collision in the year 2000 will be on all possible futures and (1) will now be true. These conditions might seem to have the effect of equating truth about the future with inevitability. This may well be the case: the conditions *are* strict, but probably no more so than everyday life ones for the truth of future-tense statements. The man who appears in the turbine-house and asks breathlessly if it is true that the dam is going to burst is not inquiring whether the bursting is a feature of only *some* possible futures. [McCall (1976) [Tooley (1999), 280]]

Now, the kind of theory McCall is advocating is a branching time theory. In addition, it is one of a dynamic variety, indicating that the complete state-description of the universe (the universe tree) is different at different times.[51] According to McCall, "each moment of time [...] defines a separation of the universe into past and future that is ontological rather than epistemological". [*ibid.*, 263] Assuming this model to depict the time flow in the objective sense (for the sake of an argument), we end up with another possible answer to the question of the qualitative difference between past and present. If, on the basis of the account given by McCall, we live in a universe with branching time, these branches

[50] The **Law of Bivalence** says essentially: For any proposition P, P is either true or false.

[51] In terms of the universe tree representations, the difference amounts to this: For all pairs of the parameters t_1 and $t_2 \in \mathbb{R}$ (i.e. 'values' of time), if $t_1 < t_2$, then the universe-tree at t_2 is a proper subtree of the universe-tree at t_1. [McCall (1976) [Tooley (1999), 263]]

and the processes forming them define the separation into past and future. But, these separations depend on the choice of the frame of reference. It follows that given an event E, there exist pairs of observers and frames $< O_i, F_i >$ such that for some observers in particular frames the event should have happened which for the other observers in different frames merely lie on possible future branches. This makes the distinction between past and future frame-dependent rather than mind-dependent.

An *intermezzo:* The metaphysics of determinism

Although Carnap was not explicitly concerned with the time-honoured problem of determinism, he could not avoid taking implicitly a stance with respect to it. His occupation with the foundations of geometry and their relation to physics steered him onto the vestiges of deep ontological problems. The very formulations of space-time-topologies with which he was concerned in the 1920s implied definite philosophical commitments as to the ontological and epistemological reverberations of determinism in physics. To make these commitments explicit, it is necessary to adduce the outlines of the conceptual difficulties and complexities inherent in the metaphysics of determinism. I will therefore finish this section by presenting the main issues concerning determinism in its relation with physics. Here, I will draw heavily on the work of John Earman who has done more than any other contemporary philosopher of science to clarify these issues from the point of view of physics. This final section serves simultaneously as a presentation of the ramifications of the constitutional program that Carnap was involved in. However, it is implausible that Carnap could have foreseen all of the difficulties involved.

As is well known, in pre-General-Relativity-Theory ("General Relativity Theory" is henceforth abbreviated as GTR) physics, space(time) was thought to serve as a fixed background against which the drama of physics is enacted [Earman (2007)], its *dramatis personæ* consisting of classical fields (gravitational, electro-magnetic etc.) and particles. Furthermore, in pre-Quantum-Mechanics ("Quantum Mechanics" is henceforth abbreviated as QM) it was in addition assumed that there exists a unique set $\mathcal{O}$ of genuine physical magnitudes[52] each of which takes a determinate value at every moment of time [*ibid.*]. What, then, distinguishes genuine physical magnitudes from other physical magnitudes? The other magnitudes that are of concern for a physicist are usually described as dispositional, but this kind of a distinction seems quite vague.[53] Nevertheless, the logical relation between dispositional and non-dispositional magnitudes was assumed to be

[52]In the classical philosophical parlance, these are also referred to as "observables".

[53]It is, as a matter of fact, quite challenging to give a satisfactory account of dispositional qualities and quantities. The simplest analysis of dispositions is often given along the following lines: **SCA** (Simple Conditional Analysis): An object is disposed to M when C iff it would M if it were the case that C. But, as is easily seen, this does not amount to an adequate criterion for differentiating between genuine and dispositional *physical* magnitudes. Another, more sophisticated, but nevertheless, defective characterization of a dispositional property is the following: **Sophisticated** An object is disposed to M when C iff it has an intrinsic property B such that, if it were that C, and if the object were to retain B, then the object would M because C and because it has B. [SEP]

captured essentially by insisting that the former supervene on the latter.[54] To adduce the classical picture of determinism in physics *ab initio* we shall have recourse to a couple of additional auxiliary concepts. A *history* H is a map from $\mathbb{R}$ to tuples of values of the basic magnitudes, where for any $t \in \mathbb{R}$ the *state* $H(t)$ gives a snapshot of behaviour of the basic magnitudes at time t. The world is *Laplacian deterministic* with respect to $\mathcal{O}$ just in case for any pair of histories, H_1, H_2 satisfying the laws of physics, if $H_1(t) = H_2(t)$ for some t, then $H_1(t) = H_2(t)$ for all t. The brevity of these stipulations obscure the scope of the scrutiny that they deserve. I will now elaborate on this on a case by case basis.

¶It is essential to note that the parameter 't' appearing here in the definition of histories is a *global time function*. A definition is as follows: a global time function is a smooth map $t : \mathcal{M} \to \mathbb{R}$, where $\mathcal{M}$ is the spacetime manifold, such that for any $p, q \in \mathcal{M}$, $t(p) < t(q)$ just in case there is a future directed timelike curve from p to q.[55] The second requirement concerning the orientation of time presupposes that a solution to the problem of the direction of time has been found. [*ibid.*, 1370] It is to be kept in mind that the definition of the global time function applies equally to classical, special relativistic and general relativistic spacetimes. Global time functions are most easily definable in classical physics, where all spacetimes are characterized by the notion of absolute simultaneity, that is, simultaneity independent of all observers. A timelike curve is then specified to be one which is oblique to the planes of simultaneity; thus, in classical setting t is defined up to a transformation of the form $t \to t' = t'(t)$. The case in special relativity is equally straightforward; timelike curves are defined to be those whose tangents at any point lie

[54]The notion of supervenience is at least as ambiguous as that of disposition: different definitions abound. One of the most natural varieties with respect to physical magnitudes (which are our concern here) is perhaps given by the **Strong Individual Supervenience** scheme: A strongly supervenes on B iff necessarily, if anything x has some property F in A, then there is at least one property G in B such that x has G, and necessarily everything that has G has F; or, in modal-operator formalism: $\Box\forall x \forall F \in A[Fx \to \exists G \in B(Gx \& \Box\forall y(Gy \to Fy))]$. In general, the weakest form of supervenience can be stated by stipulating that X supervenes on Y if and only if within the range of possible cases, there is no difference in X without a difference in Y. Now, according to this scheme, the strength and type of the supervenience relation is dictated by the space of possible cases. How this space is determined, depends on the context. Here, as we are dealing with physical systems, the range of possible cases is determined by the laws of physics, i.e. the possible cases in question are all those cases c that are compatible with the physical laws.

[55]A few more technical conditions are necessary for these claims to hold: firstly, it is required that the spacetime is temporally orientable and secondly, that one of the orientations has been singled out as giving the "future" direction of time. (A relativistic spacetime (M, g_{ab}) is temporally orientable iff there exists a continuous everywhere defined timelike vector field on M. If such a field exists, reversing the arrows gives another such field. The choice of one of these fields as "pointing the way to the future" is what is meant by the assignment of a time orientation.) N.B. A general relativistic spacetime may not be temporally orientable, but a covering spacetime always is. (A (double) covering spacetime of a spacetime (M, g_{ab}) may be defined as the set of all pairs (p, α) where $p \in M$ and α encodes one of the two temporal orientations at p. A projection

$$\pi : (p, \alpha) \to p$$

maps events of the covering spacetime back into M. Locally, this projection must be a diffeomorphism in order for the covering spacetime to inherit the local differential structure of M (a diffeomorphism is an isomorphism of smooth manifolds). In case the double covering spacetime consists of two disparate parts, there exist two different, globally consistent assignments of a temporal orientation and (M, g_{ab}) is time-orientable. If the covering spacetime is connected, there exist smooth transitions from one time orientation to the other and vice versa. This implies that (M, g_{ab}) is not time-orientable, but that the covering spacetime is). All these results are rigorously derived in [Hawking & Ellis (1973)], section 6.1.

inside the light cone in that point. General relativity provides instances of some pathological behaviour, the most famous of which are without doubt the Gödelian spacetimes. In *Gödel spacetimes* there can be no global time function, and the paradigmatic case of global Laplacian determinism as defined above is undermined. However, in case there exists a global time function for relativistic spacetime, this guarantees that many then exist. The choice of this global time function is essential, because the success or failure of Laplacian determinism depends on it; a poor choice of a global time function banishes the Laplacian Demon from the relativistic quarters of spacetime.

¶The second issue of concern revolves around the distinction between natural laws and initial or boundary conditions that the Laplacian formulation of determinism assumes. Whether the form of determinism advocated is cogent or not depends largely upon how clearly the two have been demarcated from each other and how their relative roles have been defined. The philosophically oriented literature on natural laws is formidable. However, most of it does not seem to touch, at least in an illuminating way, the problems concerning the practice of physics. The quite simplistic approach that most works in this vein take with respect to physics is apt to deteriorate one's interest in the discussions entertained in them. However, the stance that is taken on the other side of the fence, that is, among the scientifically oriented philosophers, is often equally bale with respect to the pragmatic and overtly simplistic attitude in setting the framework for discussion. For example, John Earman proceeds in his otherwise meritorious account quite straightforwardly by merely stipulating that "an acceptable account of laws must satisfy the empiricist constraint that the laws supervene on the totality of non-modal, particular facts".[56] [*ibid.*, 1371] Although Earman makes a justifiably sarcastic comment on philosophers' tendency to speculate on non-empiricist laws, the suggestion of his to model his account of physical laws in terms of the account of David Lewis seems philosophically unfinished, to say the least: "the laws of physics are the axioms or postulates that appear in *the ideal theory of physics*[57]: the laws of physics are the axioms or postulates that appear in the ideal theory of physics, where the ideal theory is the one that, within the class of true theories, achieves the best balance between simplicity and information content.[58]" Earman punctuates his thesis with the note that all of the relevant candidates for deterministic laws in the history of physics involve only a relatively small subset $\mathcal{B} \subset \mathcal{O}$ of basic occurrent magnitudes, with the assumption that the remaining ones supervene on those of $\mathcal{B}$. A paradigmatic example of these is classical particle mechanics where the basic "occurrent magnitudes" are the positions and momenta of particles, that is $\mathcal{B} = \{(q_i^1, q_i^2, q_i^3, p_i^1, p_i^2, p_i^3) | i \in A \subset \mathbb{N}\}$, where A denotes an ordered and indexed subset of $\mathbb{N}$ ($A = [k] = \{1, 2, 3, \dots, k\}$). Here Earman is in good company. The assumption (or ideal) of the simplicity of the true physical laws has been a mainstay among heuristic principles of physicists in the last century. Especially Hermann Weyl shared this conviction with Einstein (that simplicity must figure into the account of laws), but in a certain sense he must have been dimly mistrustful of it for he went on to state that: "this circumstance is apt to weaken the metaphysical power of determinism, since it makes the

[56]This weak form of the empiricist constraint is coined by David Lewis as the condition of "Humean supervenience".
[57]Sic! My emphasis.
[58]Notice the resemblance to Carnap's aspirations in his early work in the 1920s.

meaning of natural law depend on the fluctuating distinction between simple and complicated functions or classes of functions". [Weyl (1932), 42.] This reflects, perhaps, the modest but clear-sighted wisdom of a mathematician. Be that as it may, the issue is not settled, and to pursue the investigation further may be easier for someone fortified by faith than someone waiting to be convinced by argument.

¶The third issue hinges on the prospects of envisaging the world as only partially deterministic. More exactly, this is tantamount to the following scenario: the world could be deterministic with respect to partial histories defined by the values of magnitudes in some proper subset $\mathcal{D} \subset \mathcal{O}$ of the occurrent physical magnitudes but non-deterministic with respect to partial histories defined by the values of magnitudes in some other proper subset $\mathcal{N} \subset \mathcal{O}$. (ibid., 1371) The plausibility of this scenario is best assessed by considering two cases: (i) both $\mathcal{D}$ and $\mathcal{N}$ are basic magnitudes; (ii) one or the other of the subsets comprises non-basic magnitudes. First (i). This scenario comes down to this: assume that the subset $\mathcal{N}$ denotes the set of quantities undergoing a nondeterministic evolution. Then it is necessary that the evolution of the quantities in $\mathcal{D}$ is unhampered by the evolution of the magnitudes in $\mathcal{N}$, that is, the magnitudes in $\mathcal{N}$ must not interact with those in $\mathcal{D}$. Alternatively, there would have to be a conspiracy of sorts, where the disturbing effects of $\mathcal{N}$ cancel out, which amounts to the same thing operationally. It is therefore highly dubious whether such a scenario could really work. Second, the possibility stated in (ii) gives rise to an even more recalcitrant problem; in particular, in the context of certain physical systems, stochastic processes on one level can supervene on deterministic processes at a lower level. Thus, the phenomenally based inference from observed stochastic behaviour to indeterminism is in general, as is said, fraught with peril.[59]

¶The fourth issue is of a mathematical nature. Because the laws of physics — in the most typical cases at least — take the form of differential equations, the thesis of Laplacian determinism raises the question as to the possibility of giving an *initial value formulation* for the particular problem at hand. This formulation is tantamount to the question whether for arbitrary initial conditions there exists a unique solution of the differential equation agreeing with the given initial data.[60] The nature of the initial data are in the mathematical sense clear-cut: they comprise the instantaneous values of the independent variables in the equation together with the instantaneous values of a finite number of time derivatives of these variables. The locution "arbitrary" is not to be taken too liberally, however; the nature of the initial conditions depends crucially on the equations of motion, that might impose non-trivial conditions on the initial data.

¶The fifth issue elaborates on the point made in the last paragraph: in relativistic setting, the field equations can often be solved by separating them into constraint equations and evolution equations. The former place restrictions on the initial data, and the latter govern how the initial data satisfying the constraints evolve over time. The most conspicuous examples of these types of factoring field equations are provided by Maxwell's equations for electromagnetism ad Einstein's gravitational field equations. Here the characteristic

[59] That is, it is not *universally valid*.

[60] More rigorous conditions could also be imposed on the solution, such as determining, for example, whether the solution depends continuously on the initial data.

property of the solutions is that once the constraint equations are satisfied they continue to be satisfied over time (this is guaranteed by the evolution equations). This consequence of the separability of the constraint and evolution equations marks a distinctive feature of deterministic equations.

¶Finally, the sixth issue is more philosophical. Earman starts by pointing out that "while there is no *a priori* guarantee that the laws of the ideal theory of physics will be deterministic, the history of physics shows that determinism is taken to be what might be termed as a 'defeasible methodological imperative'; start by assuming that determinism is true; if the candidate laws so far are not deterministic, then presume that there are other laws to be discovered, or that the ones that are so far discovered are only approximations to the correct laws; only after long and repeated failure may we entertain the hypothesis that the failure to find deterministic laws does not represent a lack of imagination on our part but reflects the fact that Nature is non-deterministic". [*ibid.*, 1372] There is an interesting tension between this statement and the second issue above, namely, the insistence on greatest possible simplicity and informational content even at the expense of the program of determinism. What is at stake here is a measuring of a balance between distinct *heuristic principles*. Traditionally, all of the above mentioned heuristic maxims have played a prominent role in the development of physical theories; in some notable cases they have even worked in tandem. If the clear-sighted observation of Weyl is to be taken seriously, the heuristic principles used by physicists should certainly merit further scrutiny. But this is a task to which Carnap's method of explication is ideally suited. I will investigate the issue of determinism in its relationship to the concept of randomness from precisely this point of view in Chapter 6.

4.4 Russell and *Our Knowledge of the External World*

For the generation of young philosophers and scientists that were educated around the period spanning the World War I, Bertrand Russell signified a hero and role model, whose penetrating insights into the interaction between logic, mathematics and philosophy had opened up a completely new approach to philosophical enquiry as an intellectual undertaking. It is commonplace to hear statements that Russell introduced what is known as philosophical analysis. This is misleading in two senses. First, Russell was not the first to introduce 'analytic' tools into philosophy. Second, the term "analysis" is ambiguous; it is not at all clear what the method of analysis in philosophy amounts to (On the ambiguity of the term "analysis", see Chapter 3.). As we have seen, 'analytic' approaches had been common in philosophy ever since Kant presented the analytic/synthetic distinction as the overarching theme around which the most interesting and challenging philosophical questions were winded. Indeed, much of the momentum of the discussions about the foundations of mathematics in the late nineteenth century were in large part due to the Kantian manner of approaching the problem, at least in its preliminary formulations. Thus Frege's position as the 'grandfather' of analytic philosophy stemmed from the German tradition in logic where the dispute (at the level of conflicting formulations at least) between Kant and Bolzano marked its starting point. Prominent German logicians in this lineage include at least Hermann Grassmann (1809–1877) who was philosophically influenced by the neo-Kantian Friedrich Schleiermacher (1768–1834), and who published in 1844 an influential book *Die Lineale Ausdehnungslehre*, effectively a pioneering work on linear algebra; Hermann's brother Robert Grassmann (1815–1901), a philosopher and logician by training, and a teacher and publisher by profession, whose *Die Formenlehre oder Matematik* was a landmark in logic; and Ernst Schröder (1841–1902), a brilliant logician following traditional logic and C.S. Peirce in particular, his thought ruminating even over Cantorian themes in the late 1800s. The modern period of mathematical logic starts with Frege, Peano, Whitehead and Russell. This transition is one of the most puzzling in the history of logic since the new developments largely overshadowed the work of the logicians in the German tradition, notably the work of Schröder, whose contributions included many novel insights and partly overlapped with the important discoveries of C.S. Peirce. Analytic philosophy was, of course, to be influenced mainly by the 'new' mathematical logic whose main exponents were Frege and Russell. Frege's contributions to the analytic methodology of philosophy we have briefly discussed above. Now it is time to make explicit those contributions of Russell that played a decisive role in influencing the thought of Carnap.

There is a well-known passage in Carnap's *Autobiography* where he acknowledged the tremendous significance of Bertrand Russell on his thinking. The latter's *Our Knowledge of the External World*, in particular, worked as a catalyst for Carnap's still dormant but tremendous powers of discernment and half-articulated views on the significance of philosophy and its connection with science. As we saw in the section **4.1.4.**, the main motivation for Carnap's philosophical work sprang from the idea of developing a comprehensive system of knowledge. Although he was constantly occupied with various special problems in the intersection of philosophy and science before 1922, these investigations were always subjected to, and put in a perspective by, the overall aim of general cognitive systematization: "I worked on many special problems, always looking for new approaches and improved solutions. But in the background there was always the ultimate aim of the total system of all concepts. I believed that it should be possible, in principle, to give a logical reconstruction of the total system of the world as we know it."[61] [UCLA 1957a, E4] This is the intellectual situation in which Carnap found himself at the time of his dissertation *Der Raum*. The leading theme of the *Aufbau* is already there, but the *concreta* of its realization remain still in the dark. André Carus has found compelling evidence for maintaining that Carnap advocated this line of work in a sustained and coherent manner all the way up to the publication of the *Aufbau*. The evidence stems from the numerous notes that Carnap wrote during this period. According to Carus, even in the earliest of these notes Carnap "implicitly rejects both the Marburg and the phenomenological responses to Helmholtz's empirical investigation of perception." [Carus (2007), 139] The particular quotation from the notes (dated August 1920 and headed "Skeleton of Epistemology" [*Skelett der Erkenntnistheorie*] reads:

> The first given: experiences (facts of consciousness) . . . In some cases I can observe, rather than experience itself, a particular aspect of the experience:

[61]Cited in [Carus (2007), 139].

its 'object'. (So not 'intentional relation'); these experiences we call 'ideas [*Vorstellungen*]'. Some ideas are particularly singled out (give the criterion!): '*sensations* [*Empfindungen*]'.[62] [ASP 1920c]

Contrary to the conclusion of Carus, I am not ready to downplay the role of phenomenology in Carnap's thought, especially during this early phase of his thinking. Although the passage quoted hints at a possible rejection, claiming that such a rejection has effectively taken place is unjustified. After all, Carnap uses a phrase "in some cases ...". As these notes stem from a very early period, there is no plausible case for attributing to Carnap a view that is hostile to phenomenology. Indeed, much of the underlying motivation and conceptual machinery of *Der Raum* (1922) stems from Husserlian thought. The role of phenomenology is also visible in the *Aufbau* albeit not anymore as strong as in *Der Raum*. Questions about Husserlian influence notwithstanding, it is clear how Carnap's thought here exercises itself with a conspicuously Russellian idea, viz. that of a progressive series of (logical) constructs founded on the immediately given sensations, or sense-data, that amounts to an instance of an application of the professed general method of constructing a corpus of knowledge on a strictly empirical basis. In the material investigated by Carus there is a further set of notes entitled "Analysis of the World Picture [*Analyse der Weltbildes*]" where this issue is put in a succinct phrase: "*The point of knowledge: bringing order to the chaos* of sensations." This process of knowledge acquisition was to be built on a "progressive hierarchy of ordering [*Stufengang des Ordnens*]" from groups or patches "in the momentary visual field [*im Augenblicksgesichtsfeld*]" to enduring physical objects to the ultimate end-point "regularity of succession; *nature*." Moreover, Carnap seemed to have thought that the third stage pertaining to the discovery of the regularities in nature, or laws of nature were effectively vehicles for extrapolating the scope of our knowledge to future: "Every physical law is basically a claim about expected sensations!" But rather than advocating a distinctively Machian point of view, Carnap went on to stress that the theoretical priority in science did not lay in the given but in the logical constructions to the extent that "[*Retrospective shift of viewpoint*]: The constructions now go by the name 'real world', the immediate by 'just appearances' (and sometimes 'illusions')."[63] [ASP 1921a] Instigated by these general considerations Carnap attempted to elaborate the general scheme of knowledge acquisition he had envisaged by theoretically solidifying it. This necessitated a logical criterion for deciding which elementary sensations were similar in their content. In devising a notation for indicating "identity of content [Inhalts identität]" between these 'elementary experiences' – a veritable premonition with respect to the essential ideas of the *Aufbau* – Carnap made his idealistic conception of knowledge (which he held at the time) more precise. The interpretation of the relation of content as identity of content was, however, an added "rational ingredient [*rationale Zutat*]." Thus, Carnap is here concerned with essentially neo-Kantian questions about deciding which components of our knowledge are "immediate [*unmittelbar*]" and which "rational ingredients". Carus reports that in the notes headed "On the analysis of experiences [*Über die Analyse von Erlebnissen*]", "Carnap searches for a criterion to determine which relations or distinctions among experiences and classes of experiences

[62] Cited in [Carus (2007), 139].

[63] Quoted passages cited in [Carus (2007), 139–140].

are genuinely and unquestionably 'immediate' or 'experiental [*erlebnismäßig*]'." [Carus (2007), 140] As we have seen, in the context of the Marburgian interpretation of the *a priori* constituents of our knowledge, such an attempt at distinguishing and identifying the two elements of (empirical) knowledge would have been meaningless because according to that interpretation experience generally was permeated with such rational ingredients and would essentially be impossible without a rational basis. Carnap was thus following in this issue the scientifically oriented neo-Kantians such as Helmholtz and individualists like Dingler,[64] advocating a view that such a demarcation could indeed be validly made. As a matter of fact, Carnap's dissertation about the foundations of geometry (that formed the basis for the 1922 *Der Raum*) was effectively built on the presupposition that it can be unambiguously decided "which a priori and which empirical elements turn up in each of these spatial systems."[65] [UCLA 1920a, 37] This particular manner of conceiving the analytic-synthetic distinction, with special emphasis on the pragmatic considerations pertaining to it, was to remain one of the most important basic presuppositions of Carnap's philosophy all the way up to the latest discoveries in inductive logic.

Now, a most forceful impulse to Carnap's thinking along the lines of 'construction theory' alluded to above came from reading Russell's book *Our Knowledge of the External World – as a Field for Scientific Method in Philosophy* in early 1922. Its influence on Carnap stemmed perhaps equally from the charming rhetoric of Russell as well as from its compelling introduction of a general agenda for future philosophy as a whole. Stylistically Russell's writings are, indeed, nothing short of brilliant, representing in philosophy, what Virginia Woolf's essays represent in literary studies, "aristocrats of essays [. . .] witty, beautifully mannered and mellow."[66] In the opening chapter the new conceptual riches made available for philosophers were alluded to in a programmatic tone:

> Philosophy, from earliest times, has made greater claims, and achieved fewer results, than any other branch of learning. Ever since Thales said that all is water, philosophers have been ready with glib assertions about the sum-total of things; and equally glib denials have come from other philosophers ever since Thales was contradicted by Anaximander. I believe that the time has now arrived when this unsatisfactory state of things can be brought to an end. [Russell (1914a), 13]

Even if the promise might have sounded portentous, Russell gave ample support for such a sweeping claim. Indeed, the book tried to propound the idea that philosophy should be founded on conceptual or logical analysis. Such an analysis professed to be based on the newly constructed conceptual tools of mathematical logic, particularly the

[64]Called a "radical conventionalist" by Carnap, essentially holding similar views as Poincaré in the sense that the background assumptions and presuppositions of geometry and physics cannot be extracted from experience, nor derived by any form of transcendental deduction. He diverges from Poincaré in that he does not believe that there is freedom to choose between alternative assumptions. Characteristic of Dingler's position is his most uncompromising adherence to the classical edifices of Euclidean geometry and Newtonian mechanics (Dingler notoriously opposed Einstein's theory of relativity).

[65]Cited in [Carus (2007), 140].

[66]Rumer Godden in a review of Woolf's *Granite and Rainbow* in NY Herald Tribune.

theory of relations. What, then, did this analysis consist in? Russell did not settle with merely a vague and general statement of this method, but attempted to make it as precise as possible. Hence, he elaborated heavily on this idea in the body of the book:

> In every philosophical problem, our investigation starts from what may be called 'data', by which I mean matters of common knowledge, vague, complex, inexact, as common knowledge always is, but yet somehow commanding our assent as on the whole and in some interpretation pretty certainly true. . . . We are quite willing to admit that there may be errors of detail in this knowledge, but we believe them to be discoverable and corrigible by the methods which have given rise to our beliefs, and we do not, as practical men, entertain for a moment the hypothesis that the whole edifice may be built on insecure foundations. In the main, therefore, and without absolute dogmatism as to this or that special portion, we may accept this mass of common knowledge as affording data for our philosophical analysis. It may be said — and this is an objection which must be met at the outset — that it is the duty of the philosopher to call in question the admittedly fallible beliefs of daily life, and to replace them by something more solid and irrefragable. In a sense this is true, and in a sense it is effected in the course of analysis. But in another sense, and a very important one, it is quite impossible. While admitting that doubt is possible with regard to all our common knowledge, we must nevertheless accept that knowledge in the main if philosophy is to be possible at all. There is not any superfine brand of knowledge, obtainable by the philosopher, which can give us a standpoint from which to criticise the whole of the knowledge of daily life. The most that can be done is to examine and purify our common knowledge by an internal scrutiny, assuming tha canons by which it has been obtained, and applying them with more care and with more precision. Philosophy cannot boast of having achieved such a degree of certainty that it can have authority to condemn the facts of experience and the laws of science. The philosophic scrutiny, therefore, though sceptical in regard to every detail, is not sceptical as regards the whole. That is to say, its criticism of details will only be based upon their relation to other details, not upon some external criterion which can be applied to all the details equally. The reason for this abstention from a universal criticism is not any dogmatic confidence, but its exact opposite; it is not that common knowledge *must* be true, but that we possess no radically different kind of knowledge derived from some other source. Universal skepticism, though logically irrefutable, is practically barren; it can only, therefore, give a certain flavour of hesitancy to our beliefs, and cannot be used to substitute other beliefs for them. [*ibid.*, 72–74]

Russell's attempt here to delineate the main ingredients of his analytical method struck a tone in Carnap's mind. Indeed, many of the leading themes of Carnap's mature philosophy are featured here, and it was thus not a mere sentimental feeling of indebtedness that made Carnap so openly relate the emotions that reading Russell's book aroused in him. The famous passage of the *Autobiography* is worth quoting in full:

> Whereas Frege had the strongest influence on me in the fields of logic and semantics, in my philosophical thinking in general I learned most from Bertrand Russell. In the winter of 1921[–1922] I read his book *Our Knowledge of the External World, as a Field for Scientific Method in Philosophy*. Some passages made an especially vivid impression on me because they formulated clearly and explicitly a view of the aim and method of philosophy which I had implicitly held for some time. In the Preface he speaks about "the logical-analytic method of philosophy" and refers to Frege's work as the first complete example of this method. And on the very last pages of the book he gives a summarizing characterization of this philosophical method in the following words:
>
> > The study of logic becomes the central study in philosophy: it gives the method of research in philosophy, just as mathematics gives the method in physics . . .
> > All this supposed knowledge in the traditional systems must be swept away, and a new beginning must be made. . . . To the large and still growing body of men engaged in the pursuit of science, . . . the new method, successful already in such time-honored problems as number, infinity, continuity, space and time, should make an appeal which the older methods have wholly failed to make. . . . The one and only condition, I believe, which is necessary in order to secure for philosophy in the near future an achievement surpassing all that has hitherto been accomplished by philosophers, is the creation of a school of men with scientific training and philosophical interests, unhampered by the traditions of the past, and not misled by the literary methods of those who copy the ancients in all except their merits.
>
> I felt as if this appeal had been directed to me personally. To work in this spirit would be my task from now on! And indeed henceforth the application of the new logical instrument for the purposes of analyzing scientific concepts and of clarifying philosophical problems has been the essential aim of my philosophical activity. [Carnap (1963), 13]

What especially struck Carnap in view of the problems he was engaged with at the time, was Russell's suggestion of a strategy for the task of logical construction. In a chapter entitled "The World of Physics and the World of Sense", Russell went on to draw the outlines of such a task, beginning from a remark that "Among the objections to the reality of the objects of sense, there is one which is derived from the apparent difference between matter as it appears in physics and things as they appear in sensation". [Russell (1914), 106] Russell explicitly admitted of not having attempted to actually devise a logical construction in all its completeness and detail. Rather, he quite modestly stated what was the only conceivable goal of his discussion: "It is therefore necessary to find some way of bridging the gulf between the world of physics and the world of sense, and it is this problem that will occupy us in the present lecture. Physicists appear to be unconscious of the gulf, while psychologists, who are conscious of it, have not the mathematical knowledge required for spanning it. The problem is difficult, and I do not know its solution in detail. *All that I can hope to do is to make the problem felt, and to indicate the kind of methods by which a solution is to be sought.*"[67] [*ibid.*] The sketch Russell proceeds to draw displays many of the central ideas that Carnap's later elaborations in the *Aufbau* hinged on. A basic presupposition of Russell is the impermanence of immediate data (of experience): "[i]n the world of immediate data nothing is permanent". [*ibid.*, 109] In contrast to these he defines a thing "as a certain series of appearances, connected with each other by continuity and by certain causal laws". [*ibid.*, 111] The construction is alleged to be capable of proceeding in using only the logical notions of relations and classes. Of course, this was a major ingredient in the approach's attractiveness for Carnap. If the Frege-Russell logic could effectively be applied for this purpose, then there inescapably opened up new horizons for applying the same tools for constructing the entire edifice of knowledge *deductively*, without recourse to more classical conceptual tools of philosophy, including Spinoza's *scientia intuitiva*, Leibniz's *characteristica universalis* or Kant's transcendental logic. An overall cognitive systematization of knowledge and concepts could be acquired far more economically and efficiently than had been dreamed of. There are, however, also problems within the account of Russell that are not compatible with the stated purpose of the construction. Russell insists that he draws on the maxim "which inspires all scientific philosophizing, namely 'Occam's razor': *Entities are not to be multiplied without necessity*." [*ibid.*] This gives a markedly ontological flavour for Russell's discussion, amounting to an inconsistency since the main purpose of the whole discussion is to *construct* physical objects out of the given (i.e. sense-data). Russell, although having discarded talk about permanent things, has, nevertheless, retained an ontological commitment to regularities such as continuity and causal laws, and even goes so far as to define things at the outset as connected series of appearances. Now, the 1921–1922 Carnap would certainly have found such commitments unacceptable. There arise, then, two important questions with respect to Carnap's connection with Russell's book, which Carus has rightly brought forth: (1) Why did Russell's preoccupation with ontology not prevent Carnap from enthusiastically endorsing the "logical-analytic method of philosophy"? (2) Why did his acceptance of this programme nonetheless harden Carnap's rejection of ontological questions as a legitimate object of discourse? [Carus (2007), 143] The answer to the first question lies largely in weighing of the relative merits of the application of the "logical-analytic method". Even if Russell retained in part a conspicuously ontological vocabulary in tackling the problem of logical reconstruction, there were many benefits to be acquired from the use of the logical method. Indeed, Russell had severely criticized the *a priori* character of philosophical reasoning in the 'classical tradition' of philosophy:

> The original impulse out of which the classical tradition developed was the naïve faith of the Greek philosophers in the omnipotence of reasoning. The discovery of geometry had intoxicated them, and its *a priori* deductive method appeared capable of universal application. They would prove, for instance, that all reality is one, that there is no such thing as change, that the world

[67] Italics added.

> of sense is a world of mere illusion; and the strangeness of their results gave them no qualms because they believed in the correctness of their reasoning. Thus it came to be thought that by mere thinking the most surprising and important truths concerning the whole reality could be established with a certainty which no contrary observations could shake. As the vital impulse of the early philosophers died away, its place was taken by authority and tradition, reinforced, in the Middle Ages and almost to our own day, by systematic theology. Modern philosophy, from Descartes onwards, though not bound by authority like that of Middle Ages, still accepted more or less uncritically the Aristotelean logic. Moreover, it still believed, except in Great Britain, that *a priori* reasoning could reveal otherwise undiscoverable secrets about the universe, and could prove reality to be quite different from what, to direct observation, it appears to be. It is this belief, rather than any particular tenets resulting from it, that I regard as the distinguishing characteristic of the classical tradition, and as hitherto the main obstacle to a scientific attitude in philosophy. [Russell (1914a), 15–16]

Although Carnap was definitely a philosopher of an anti-ontological persuasion in 1922, partly through the influence of some prominent philosopher-scientists as Helmholtz and Dingler, he could not avoid being affected by the impression of a radical shift in the underpinnings of philosophical truth-seeking that Russell reported to have taken place. As matters stood, the newly discovered logical tools promised to bring within the grasp of philosophers solutions of problems that the more traditional methods of philosophy had failed to deliver. Indeed, it was just because the new methods brought about such a general feeling of liberation that made Carnap and others endorse the new logical-analytic tools enthusiastically. To solidify his case for the genuinely creative task of logic, Russell had claimed that the use of logic in the classical tradition had been restricted to a negative role: "In that tradition, logic becomes constructive through negation. Where a number of alternatives seem, at first sight, to be equally possible, logic is made to condemn all of them except one, and that is then pronounced to be realized in the actual world." [*ibid.*, 18] In contradistinction, the sense of liberation brought about by modern philosophical logic was succinctly put by Russell:

> The true function of logic is, in my opinion, exactly the opposite of this [traditional function of logic]. As applied to matters of experience, it is analytic rather than constructive; taken *a priori*, it shows the possibility of hitherto unsuspected alternatives more often than the impossibility of alternatives which seemed *prima facie* possible. Thus, while it liberates imagination as to what the world *may* be, it refuses to legislate as to what the world *is*. [*ibid.*, 18–19]

It can be fairly said, then, that Carnap very much appreciated this vision that accompanied the Russell's conception of applications of logic in philosophy. It formed a background to what would later constitute the gist of Carnap's own endeavors in philosophy – the mapping out of various possibilities for concept formulation and theory construction. Indeed, this aspect of Carnap's thought has given rise to novel characterizations of which that by Thomas Mormann is the most illuminating. Mormann has advocated the view that Carnap's philosophy should essentially be seen as a kind of 'science of possibilities' or "*Möglichkeitswissenschaft*": "Ich möchte behaupten, daß Carnap Philosophie als *Möglichkeitswissenschaft* konzipiert, d. h. als eine Theorie, konzeptuelle Möglichkeiten zu formulieren und auszuloten. Anstatt für einzelne Thesen zu argumentieren, eröffnet Carnap Räume konzeptuellen Möglichkeiten. Dies ist eine Invariante, die für alle Phasen seinen Denkens charakteristisch ist." [Mormann (2000), 210] [68] This aspect of Carnap's thinking becomes more and more pronounced towards the end of his career. I will revert to this later.

What is slightly harder to address is the second question why the assimilation of the logical-analytic methodology led to a more critical attitude towards ontology than before. Whereas Carnap had earlier entertained some kind of version of Helmholtzian 'ontological humility', there now seemed to arise a conspicuous disdain for ontological questions in general. How is this explained?

Carus has attempted to answer this question by emphasizing the ontological disdain in Carnap's developing view the first marks of which were already visible sometime earlier. Alluding to a passage in Carnap's *Autobiography* (original version), he argues that Carnap's ontological attitude was foremost influenced by the views of prominent physicists and neo-Kantians whose works Carnap had studied. Especially the influence of Kirchhoff, Boltzmann and Mach is acknowledged by Carnap, along that of Natorp and Cassirer. [Carus (2007), 144n5] The question remains, whether these influences alone were decisive for the gradual solidification of ontological disdain in Carnap's thought, or whether there were perhaps some additional features that could also have in part determined the intellectual attitude that Carnap finally adopted. A plausible option is that, as Carus maintains, Carnap saw his position to be unique in the sense that he was best equipped to take up the task that Russell had envisaged, both in terms of possessing adequate technical skills as well as being conscious of the (ontological) difficulties that the Russellian approach was fraught with, thereby making him capable of suggesting ways to circumvent those very difficulties. Thus, the prospects of improving the work begun by Russell might well have worked as a stimulus in solidifying the requirement of ontological parsimony. In a letter which Carnap sent to Russell along with the final version of *Aufbau* in 1928, these same themes were still very much in the foreground:

> I want to draw your attention straightaway to two points in which I have been compelled to depart from your conception. These points of difference

[68] A translation reads: "I want to maintain that Carnap conceives philosophy as *the science of possibilities*, i.e. as the theory of articulating and exploring conceptual possibilities. Rather than arguing for particular theses, Carnap opens up spaces of conceptual possibilities. This invariably characterizes all phases of his thought." Mormann makes yet a suggestion that we should conceive the Carnapian science of possibilities as closely akin to the "sense of possibilities" that figured as an important counterpart to the "sense of reality" in Robert Musil's great novel *The Man Without Qualities* [*Der Mann Ohne Eigenschaften*]. What is more important, the connection between the views of the two men go beyond mere surface similarities. Indeed, as I elaborate elsewhere in this book, there is a remarkable similarity between their attitudes to some central questions of modernism and culture at large.

> do not, however, rest on differences in basic attitude, which seems to me to be entirely shared between us. The differences arise, rather, precisely from my attempt to follow through on this basic attitude of yours more consistently than has been done so far. I therefore believe myself here to have been 'more Russellian than Russell' ["*Russellischer als Russel*"]
> I placed a motto at the beginning of the book (p. 1) which I want to call the 'construction principle'. Now I believe you have violated your own principle by not constructing but inferring the heteropsychological (see §140). I believe myself, through the way described of constituting the heteropsychological (§57f., §§140–143), to have followed this principle. In the same way I believe myself, through strict adherence to the 'autopsychological basis' ('methodological solipsism'), to have followed your principle more closely than has been done so far. You yourself call the retention of the autopsychological basis desirable, but too difficult and hardly feasible. Through my system I believe myself to have shown its feasibility, even if I have to admit that there can be no question yet of a completely satisfactory implementation of the system in all respects.
> The second point of difference concerns the realistic way of putting the question (see comments on the literature in §176). Here, too, I believe myself to have carried through your basic attitude more consistently by rejecting the (metaphysical) concept of reality. I believe that these questions about reality — and thus the entire philosophical controversy about realism — have no point or sense whatever [*überhaupt keinen Sinn*] (§§175–178 and part 2 of the pamphlet 'Pseudoproblems', which is also being sent to you).[69] [ASP 1928b]

This '*post mortem*' review of the success of his constructional undertaking in the *Aufbau* shows clearly how confidently Carnap had assimilated all the necessary issues and developments that he envisaged in defining the fundamental aims of the project in 1922. In the first place, the early neo-Kantian formulation of the three separate domains of knowledge had been given up, as a consequence of which knowledge consisted now only of two domains: the domain of pure intuition (for which an independent role within the structure of knowledge had been conferred in *Der Raum*) was disposed of. In its phenomenological form, that had been retained in the 1922 book, it ceased to exist, and was gradually replaced by a Wittgensteinian variety, envisioned as non-cognitive, being no longer articulable in language. This was surely tantamount to a total eradication of the Kantian outlook that still dominated *Der Raum* together with Husserlian elements. In the second place, a more intimate connection had been made with empirical psychology with a view that the basic questions aroused earlier (in a Vaihingerian formulation) could thus be adequately answered. The influence of Max Wertheimer's writings on Gestalt psychology was especially notable, although some authors have downplayed this influence (for example Carus), giving a more important role to the impact of Hans Vaihinger and his formulation of the Kantian problems in the *Philosophie des Als Ob*.

The conceptual framework spanned by the Kantian questions that formed the back-

[69]Cited in [Carus (2007), 144-145].

ground of Carnap's work in 1919–1921 was radically transformed under the influence of Russell's book. Carus backs this claim with ample textual evidence, including several manuscripts, dating from 1922. [Carus (2007), 146f.] To make clear the connections of Carnap's thought with both Vaihinger's account of the Kantian questions and the fundamental framework of the Gestalt theorists, it is expedient to review some of the 'billets' of concepts that Carnap had tentatively formulated in his proto-*Aufbau* notes in 1921. We have mentioned above that Carnap had attempted to define the notion of "identity of [phenomenal] content [*Inhaltsidentität*]" describing it as a concept that designates "a non-definable relation of close similarity that can only be interpreted by pointing to experiences [*durch Hinweis auf Erlebnisse*]". The (phenomenal) contents that are compared consist of that of the current experience (*vorhandenes Erlebnis*) and that of the memory trace of a previous one. [*ibid.*] Through consideration of the essential characteristics of these different modes of experience, Carnap comes to ponder the problem of how this relation of identity between contents can be experienced, i.e. to what extent it possesses what he called 'experiantiality ' [*Erlebnismäßigkeit*]. Indeed, if the professed identity between two contents could be realized only by subsequent rational interpretation [*rationale Deutung*], how could it be maintained that this relation (and others akin to it) might be *experienced*? For example, considering the relation between the experience of hearing a single tone and that of hearing a triad containing that tone, Carnap seems puzzled at how we could give criteria whether the inclusion was really experienced and not merely rationally constructed. Also, considering this and a converse case, where one distinguishes a single pitch within a triad, Carnap pondered over whether these were genuinely distinguishable in experience (*erlebnismäßige verschieden*). Now, even if Carnap felt puzzled by such cases, can we say anything with the advantage of hindsight that might contribute towards finding a solution to these questions?

A straightforward case for deciding whether the *identification* of a particular tone within a given triad is 'experientially distinguishable' from the *recognition* of that same tone within a triad is provided by the following example. If we consider a simple case of comparing the tonal *qualities* of two types of triads, (1) the major triad[70] and (2) the augmented triad,[71] we may think of a following kind of an argument to establish that the two given exemplars of musical perception are, indeed, distinguishable. First, consider the case (i): a particular tone (say, $G\sharp$) is first played singly; then a triad (say, E major) is played which contains the previously played tone as a part. The task is to identify the tone that was first played in isolation within a triad that is played consecutively. What are the gross features of this kind of a process of recognition? Now, as regards the overall character of the recognition task, we see that quite certainly the case is an example of auditory retention. We hear a single tone first and are then asked to identify that tone in a sound field of several masking tones. A well known aspect of such auditory retention tasks is the added dimension of the knowledge base against which such pitch recognition takes place. Remembered auditory knowledge of tonal systems affects how tonal events are

[70] In Western music, major triad is the chord exemplifying the basic tonality of the major scale, constructed on the first degree (called the tonic) of the scale and consisting of the root, major third and perfect fifth. In the scale of C major: C, E, G.

[71] The augmented triad is constructed on the third degree of the harmonic minor scale. In this case there is an augmented fifth between the root and the leading tone. In the key of *a harmonic minor*: C, E, $G\sharp$.

remembered. Bigand has argued that such knowledge may be largely implicit, even of a sort that is called tacit knowledge. The accumulation of such knowledge takes a lot of time and even the naïvest listeners possess such tacit schemas to the extent that makes their auditory knowledge profound. [Bigand (1993) [McAdams & Bigand (1993), 231–277]] The classical learning theory of Hebb (1949) took as its main premise that such an organized knowledge base is an inescapable property of learning by mature organisms. In a cultural background in which certain musical patterns and conventions dominate, the recognition of certain 'structured' patterns is facilitated by the built-in tacit knowledge of the listener. Therefore, given a certain tone in a context of a triad (which epitomizes one of the fundamental patterns in Western music), the retention of the single tone that is a part of that chord is comparatively straightforward. But, as psychological experiments have demonstrated, recognition is even easier, if the 'interference tones' do not form a consistent tonality with the given single tone. (Therefore, recognizing the note $B\flat$ imposed on the triad E, $G\sharp$, B should be even easier.) This example demonstrates that the 'experientiality' of recognizing a single tone within a triad (or any set of interference tones) is dependent on auditory memory and the extensive knowledge base that accompanies it. Therefore, the first example exemplifies to a large degree the rational interpretation [*rationale Deutung*] of auditory data, just as Carnap emphasized. Second, consider the case (ii): a particular triad (say, *A* major) is sounded after which we pick and choose a certain tone within it. Is the experiential dimension of this experiment exactly identical with that of the first case? It can be argued that such is not the case. Whereas in the first case we have an auditory memory item of a single tone the correspondent of which we then attempt to recognize in a given triad, the second case gives rise to a phenomenon that is completely different, although the two are temporally symmetric. In the second case, we first hear the *quality* of the given triad comprised by its interval structure. It is characteristic to this quality that every tone in it has a function that distinguishes it from the others. Comparing the major and augmented triads, we immediately perceive their qualitative difference: our attention is attracted by the characteristic quality of the fifth in the latter triad, in which it forms a conspicuous dissonance with the root. Although skill can be acquired in recognizing the different qualities of various chords and intervals (especially as the harmonic context gets more complex) the most conspicuous aspect of a chord is its interval structure. Tones that have a special function within the chord (the ones creating tonal tension for example) are more easily recognized. Therefore the recognition of a tone in a chord is entirely different from the first case where a given single tone was to be identified within a set of interference tones. To reiterate, in the first case, an auditory memory item is constructed which is compared with a set of given tones afterwards. The listener retains in his mind the memory of that single tone and matches it with one of the tones of the interference cluster in the second phase. It is therefore an act of *identification*. In the second case, however, the recognition of a tone is not so much an identification as an act of *ostensive reference*: given a harmonic structure (a triad, for example) we pick a single tone from it at our whim. The quality of the harmonic structure may determine which tones we find more 'interesting' than the others, but still it is not an identification between different instantiations of a given tone; it is discerning a constituent of a qualitative structure, i.e. designating a particular tone in a cluster of tones without an act of retention that would precede it. What we first recognize is the chord quality and then we perform a kind of an *decompositional analysis* in picking out a particular tone out of that context. Indeed, there is extensive empirical evidence that people recognize the sonorous qualities of a chord as a whole *without analysing the component tones of the chord*. For example, it has been shown by studies made by Trehub & Trainor that infants recognize *good* musical patterns from those that are *not* good (good patterns being defined as those the mastery of which requires relatively limited exposure). Hence, the infants they studied showed some ability at interval-processing that enabled them to distinguish between the major and augmented triads, respectively. [Trehub & Trainor (1993) [McAdams & Bigand (1993), 278–327]] In sum, there is no reason to think of the two cases considered by Carnap as experientially indistinguishable. As I have argued, the two differ considerably with respect to both their experiential content and the cognitive capacities involved.

Such particular questions aside, what were the prominent doctrinal features of Gestalt psychology that Carnap to a large extent adopted into his thinking about these problems? One of the most succinct and elegant formulations of the main ideas of Gestalt theory is a summary given by an eminent listener of Wertheimer's lectures on epistemology, Gabriele, Countess von Wartensleben. She published in 1914 a book extraordinarily titled *An Ideal Portrait of the Christian Personality: A Description 'sub specie psychologica'*, where the following account of Wertheimer's views was given in a footnote at the beginning of the book:

> M. Wertheimer's Gestalt theory (which has not yet appeared in print, but about which I learned in a lecture on epistemological problems which he gave at the Frankfurt academy in the summer semester of 1913, and in many private conversations) contains the following basic thoughts:
> 1. Aside from chaotic, therefore not, or not properly, apprehensible impressions, the contents of our consciousness are mostly not summative, but constitute a particular *characteristic "togetherness"*, that is, a segregated structure, often "comprehended" from an inner center, which can be different according to the nature of the ideational content [*Vorstellungsinhalt*]; e.g. an optical or acoustical, or also a dynamic or intensity center. To this the other parts of the structure are related in a hierarchical system. Such structures are to be called "Gestalten" in a precise sense.
> 2. Almost all impressions are grasped either as chaotic masses — a relatively seldom, extreme case — or as chaotic masses on the way to sharper formation, or as Gestalten. What is finally grasped are "impressions of structure" [*Gebildefassungen*]. To these belong the objects in a broad sense of the word, as well as "relational contexts" [*Beziehungszusammenhänge*]. They are something *specifically different from* and more than the summative totality of the individual components. Often the "whole" is grasped even before the individual parts enter consciousness.
> 3. The process of knowing — knowledge in a precise sense of the word [*im prägnanten Sinne*] — is very often a process of "centering" of "structuring", or of grasping that particular aspect which provides the key to an orderly whole,

> a unification of the particular individual parts that happen to be present; what results is that a structured unit emerges as a whole due to, and through, this centering. The result of just this knowledge process is a springing forth [*Herausspringen*] of the Gestalt from the "not yet formed" [*noch nicht gestaltet*]. Certain appearances [*So-Färbungen*] of the parts result from the specific total conception; parts and specific states now become "understandable" on this basis. The entity that results from the knowledge process depends is many respects not only on the object, but also on the observer. Thus there are several ways of grasping many phenomena, but generally only one can be correct: that which makes all states understandable and derivable from the central "idea" and thus gives meaning [*Sinn*] to the entire given.
> The same statements made about different entities can have completely different directions, according to the way in which they 'sit' in the entity [*drinsitzen*], e.g., whether they are nearer or further from the center. Thus, e.g., in the case of "the wall is red", "red" 'sits' differently than in the case of "blood is red" (though the logical situation becomes more complicated here). Thus something completely different is meant by a complex connection such as "drinker philosopher" [*Trinkerphilosoph*], according to whether the drinker is thought to be in the philosopher or the philosopher in the drinker. [Scheerer (1931), 84–85][72]

These features figure prominently in Carnap's thinking at the time (and ever since). What especially engages one's attention is the emphasis on the gradual ascension from an initial state of chaotic impressions which, given their particular *Vorstellungsinhalt*, constitute a "segregate structure", to a higher level of *Gebildefassungen*. Moreover, the essential characteristic of experience which, from the point of view of the *Gestalt*-theorists, is "a unification of the particular individual parts that happen to be present", is precisely constituted by "relational contexts" [*Beziehungszusammenhänge*]. As we have seen, the emphasis on *structure* is one of the hallmarks of Carnap's early conception of analysis. Later we will see that the very same notion forms the centerpiece of the explicative procedures that take place in the background framework of logic and mathematics.

Now, to revert back to our central topic in this section, we can summarize Russell's ideas on analysis in a manner that makes his influence, or the aspects of that influence, on Carnap more transparent. As we have seen, Russell's conception of analysis in *Our Knowldge of External World* was quite general, comprising the idea that analysis extends our intellectual horizons by explicating the conceptual possibilities inherent in any particular problem situation. Although Russell's notion of analysis developed somewhat over the years, and although its different aspects were emphasized in various ways during his career, Russell stated in *The History of Western Philosophy*, summing up his career, that a single method was common to all his philosophical ventures. [Russell (1946), 788-789] Paul Hager has distinguished three major characteristics that describe Russell's conception of analysis, these being repeatedly emphasized by Russell: [Hager (2003) [Griffin (2003), 312–314]]

[72]Cited in [Ash (1995), 123–124].

(i) Analysis is unlikely to be final. a) new premisses may be discovered in relation to which existing premisses are results, b) possibility of alternative sets of premisses for the same results

(ii) Analysis enlarges the domains of particular subjects. N.B. The philosophy/science distinction "is one , not in the subject matter, but in the state of mind of the investigator." [Russell (1919), 1] It remains for philosophy to move to the new frontier.

(iii) Analysis leads to premisses that are decreasingly self-evident. "...it becomes obvious that, if we are to believe in the truth of pure mathematics, it cannot be solely because we believe in the truth of the set of premisses. Some of the premisses are musch less obvious that some of their consequences, and are believed chiefly because of their consequences." [Russell (1924) [Russell (1956), 325]]

We can immediately recognize here elements and principles which played a significant role also in Carnap's thinking about analysis and its application in philosophy. In the sense of a *Möglichkeitswissenschaft*, (i) and (ii) are linked; they are both statements about the opening up of new possibilities that analysis enables us to envisage. But also (iii) played a significant part in Carnap's conception of analysis; his technical contributions to the theory of probability and induction are prominent examples of this. To sum up the results of this section, we can illustrate the characteristics of Russellian analysis by the following table of the qualities of its results and premisses (as given by [Hager (2003) [Griffin (2003), 315]]):

Results (or Data)	*Premisses*
More complex	Simpler
Relatively concrete	Abstract
Common knowledge	[The outcome of special enquiry]
Vague	Precise
Logically interdependent	Logically independent
More obvious	Less obvious
Undeniable	[Disputable]
Inexact and approximate	Definite
Indubitable	Dubitable
Puzzling	[Explanatory]
Confused	Clear
Self-evident	[Requiring justification]
Ambiguous	[Unambiguous]
Disorganised	Ordered

4.5 Reconstructing the world: *Der Logische Aufbau der Welt*

4.5.1 *Tractarian semantics*: Wittgenstein's influence on Carnap

The question of Wittgenstein's impact on Carnap's thought is entwined with both historical and systematic difficulties. In the first place, there are contested issues pertaining to the actual communications between the two men and to the question about the independence of some aspects of Carnap's thought from the ideas of Wittgenstein. Foremost among these is the idea of physicalism. In the second place, the reverberations of Wittgenstein's influence on the members of the Vienna Circle extend to cover a relatively long period of time. Thus it becomes problematic, foremost in connection with Carnap, to state once and for all the import of Wittgenstein's philosophy on his thinking. It is just because Carnap frequently changed his thinking about the variety of issues he tackled that it is so difficult to link Wittgenstein's influence on him with a single phase in his career. Nevertheless, it is evident that already in the *Aufbau* there can be discerned conspicuous marks of Wittgenstein's *Tractatus*. Moreover, Carnap saw Wittgenstein's overall conception presented in that book as providing for him a solution to a fundamental problem that he had come across earlier in connection with the first attempts at a rational reconstruction. This program that had its origins in the work pertaining to the foundations of geometry in *Der Raum* and that was properly set in motion during 1922–1923, relied essentially on philosophical considerations that laid great emphasis on the notion of "structure". The prominence of this notion derived from the philosophical writings of Helmholtz and Poincaré who had propounded the general conception that the ultimate component of knowledge that science is concerned with and the one that can be said to be the *only* component of genuine knowledge is knowledge about structures, or as Poincaré put it in 1904: "What science can attain to is not the things themselves but only the relations among the things; apart from these relations there is no knowable reality." [Poincaré (1904), XIII] This principle had been assimilated by Carnap already in 1922, for he wrote in one of his position papers that "every science is a science insofar as the study of structures [*Strukturenlehre*] is contained in it."[73] [ASP (1922d), a2] This idea was very much in the focus of the questions that the neo-Kantians, with whom Carnap was associated in his student years, addressed at the first quarter of the twentieth century. Indeed, the central issues concerned the possibility of synthetic *a priori* knowledge, and the questions about *Strukturenlehre* appeared very much to be questions of this sort. Was knowledge of structure actually synthetic? Moreover, if structural properties were purely formal ones, and hence derivable from logic, was logic itself a synthetic discipline? [*ibid.*, 185] At least some of the most prominent logicians, including Frege and Russell, seemed to think in this vein. Frege had envisaged his program of logic as a vehicle for reducing mathematics to logic, but as is well known, this did not amount to a conception of mathematics with no real content. Indeed, Frege's theory of concepts in a sense salvaged the Kantian intuitions that had earlier comprised the fundamental guarantee of the syntheticity of mathematics. Russell, on his part, propounded a version of 'logicism' that, in agreement with Frege, conceived logic as an *a priori* discipline but still accommo-

[73] Cited in [Carus (2007), 163].

dated synthetic knowledge in the form of a codification of the most general features of our world. The 'universalism' of Russell's philosophy of logic has been a contested issue recently, but perhaps one of the most interesting lines of interpretation has been to take Russell as advocating a Bolzanian conception of logic.[74] Carnap was beginning to feel increasingly uneasy with such conceptions of logic. At the time of conceiving clearly the principal ideas of rational reconstruction in 1922, he still held fast to elements of Kantianism and Husserlian phenomenology, although these were already being somewhat mitigated. At least the Kantianism displayed in *Der Raum* was already minimal, as we have seen. But the phenomenological component in his thinking would remain strong albeit tacit. It is one of the unexplained mysteries of the history of twentieth century philosophy why Carnap downplayed the influence of Husserl in his explicit reports on his own work although that influence was very clearly active within the confines of Carnap's philosophical 'mechanism'. Along with the question about the interrelation with Wittgenstein, this is one of the most fascinating and important questions to make, and I will attempt to give at least a partial answer to it in these final sections of this chapter. Against this background it seem therefore questionable that "[...] though phenomenological discernment had still played a role in the identification of the basis, in the 1922 version of the constitution system, by 1925 Carnap had set the goal of 'overcoming subjectivity: transition from material to structure"', as Carus claims. [Carus (2007), 185] In this respect we have to be attentive, not merely to what Carnap explicitly says of his goals and the methods for attaining them, but also to the implicit presuppositions that can be inferred from what he says in connection with systematic issues. This is an expedient strategy to adopt in enquiring the relation of Carnap and Husserl. I will, however, begin with adducing the most important aspects of Wittgenstein's thought that demonstrably influenced the members of Vienna Circle in general, and Carnap in particular. The natural half-way house between the first phase of interaction between Wittgenstein and the Vienna Circle and the next phase of 'smooth waters' when Wittgenstein temporarily retires from philosophy, breaking all the contacts with the Circle, is the inception of Carnap's *Aufbau*. Giving an account of the main ideas and achievements of the book occupies us for the most part of the later sections. In this connection it is also natural to take Husserl's role under discussion. After that I will say something of the historical difficulties pertaining to interpreting the relation of Carnap and Wittgenstein. This will then lead to the investigation of the logical issues that Carnap was interested in through the influence of Wittgenstein, ultimately leading to the next, crucial phase in Carnap's thinking during which he becomes aware of the insurmountable difficulties with the original program of rational reconstruction. This 'crisis' will then lead to the 'liberation' brought about the ideas presented in *Logische Syntax*. But this will be the subject of Chapter 5.

Wittgenstein's picture theory

A widely discussed topic within the Wittgensteinian literary corpus is the notion of language as a picture of the world, or, as it has been coined by Hintikka and others, "Wittgenstein's picture theory". A cornucopia of analyses of this 'theory' have been

[74] A sustained argument in this direction is presented in [Korhonen (2006)].

published, attempting to scrutinize its basic tenets, methodological underpinnings, and significance. To acquire a satisfactory conception of the influence that Wittgenstein incurred on Carnap, it of primary importance to give a balanced account of the essentials of Wittgenstein's picture theory, because it constitutes the core of the presentation of the *Tractatus Logico-Philosophicus*, the book that had a tremendous impact on the members of the Vienna Circle, Carnap included.

One of the most interesting analyses of the picture theory of Wittgenstein has been given by Jaakko Hintikka. Effectively basing his analysis in part on the pioneering account of Erik Stenius [Stenius (1960)], Hintikka professes to bring to light the inhomogeneity of the issues entwined with Wittgenstein's picture theory that usually are approached quite uncritically under a unitarian view of a 'Wittgensteinian picture theory'. Indeed, he does not grow tired of emphasizing the "several different and largely independent" issues that are involved beneath the surface of the unitary approach. Hintikka maintains that at least six different tenets underlie the "picture theory". These are:

1. An elementary proposition represents the (possible) state of affairs that it represents in virtue of being an isomorphic replica of this state of affairs.[75]
2. The totality of possible combinations of simple objects matches the totality of possible elementary propositions.
3. Each name (primitive symbol) has the same logical form (logical and categorial type) as the object it represents.
4. Elementary propositions are independent of each other.
5. All non-elementary propositions are (complex or otherwise derived) pictures of facts in the same sense as elementary propositions.
6. A part of the background of all these different theses is a sixth one. It is the thesis to the effect that that in logically correct language the logical (pictorial) forms of propositions are their *syntactical* forms. [Hintikka (1996), 22]

Hintikka refers to these theses with the following locutions (I intend to follow him in this practice): (1) Elementary propositions as pictures, (2) The mirroring thesis, (3) The categorial matching thesis, (4) The atomicity thesis, (5) Complex propositions as pictures, and (6) Pictorial form as a syntactical form, or the syntacticity thesis. [*ibid.*]

One of the most conspicuous discrepancies in the collection of the main theses of the *Tractatus* is the thesis number six. It seems to be the odd one out, having presumably little

[75]This interpretive approach is essentially due to Erik Stenius. The notion of isomorphism is adapted from group theory, and gives an accurate rendering of the intended notion of resemblance between propositions and states of affairs. Consider the following definition. Given two sets G and G', a homomorphism $f : G \to G'$ is called an **isomorphism** if there exists a homomorphism $g : G' \to G$ such that $f \circ g$ and $g \circ f$ are the identity mappings (in G' and G respectively). It is trivially verified that f is an isomorphism if and only if f is bijective. (N.B. A **homomorphism** of G into G' is a mapping $f : G \to G'$ such that $f(xy) = f(x)f(y)$ for all $x, y \in G$, and mapping the unit element of G into that of G'.)

explicative force within the overall philosophical agenda that Wittgenstein professes. It concerns primarily only a very elementary fact of propositional logic, the possibility of representing every truth-function in accordance with Sheffer's result, namely, in terms of a single propositional connective.[76] How could this result have anything of relevance to do with the general program of illustrating the pictorial relations between language and the world. Perhaps we can answer this question by addressing first a rather straightforward one: what does the proposition 6 of the *Tractatus* actually say? The answer is simple: only one operation is needed for representing every possible truth-function. This operation consists, effectively, of taking the conjunction of the negations of some given propositions. On the face of it all of this seems to be quite irrelevant to the overall purpose of the *Tractatus*. But, as Hintikka has argued (quite convincingly), the accommodation of the Proposition 6 within the system is essential, because it works as a tool for extending the picture idea of propositions from elementary to complex propositions. On a classical reading the idea of a complex proposition picturing a fact or a state of affairs in the world is incomprehensible, because it would make the Proposition 5 seem redundant. Indeed, if all propositions are unproblematically taken to picture reality, what need is their, then for the thesis that complex propositions are truth- functions of elementary propositions? But readings falling under this classification have been marred exactly in their incapacity to see the full import of the proposition 6 of the *Tractatus*. Rather than being an incoherent sidestep in the development of the argument, it is in fact an essential part of it. Or so Hintikka essentially claims. Let us see to what extent this can be justified.

Now, given Wittgenstein's definition of the general form of the truth-function, we can say something of the requirements of the picturing relations between complex propositions and the world. To reiterate, the general truth-function is of such a form that that it reduces all complex propositions to a form that consists only of conjunctions of negations of propositions (in all but the simplest cases consisting of multiple layers of such conjunctions of negations of propositions).[77] How can such a truth-functional form be interpreted as a picture? We may think of the correspondence as being constituted by consecutive iterations of the two basic operations, negation and conjunction. Therefore we have to consider what the pictorial representations of these basic operations amount to. Hintikka regards these as quite straightforward. In the first place, conjunction is just the superposition of two pictures (corresponding to two (elementary) propositions), a "conjoint complex picture". [Hintikka (1996), 40] Strictly speaking, I would considerably qualify *this* type of description, but let us accept it for now, for the sake of an argument. In the second place, the negation of a picture is not only a picture, it is the very same picture but taken with the opposite sense (with the converse polarity). This alludes to Wittgenstein's notion of *bipolarity*.[78] Hintikka describes this idea (the relation between p and $\sim p$)

[76]In Wittgenstein's own words, "[6.001] What this says is just that every proposition is a result of successive application to elementary propositions of the operation $N(\bar{\xi})$". [*Tractatus*, 6.001]

[77]For example, considering a very simple case in propositional logic, the sentence "$F = (A \vee B) \rightarrow (\neg B \wedge A)$" can be transformed first into the CNF (conjunctive normal form) "$(\neg A \vee \neg B) \wedge (\neg A \vee A) \wedge (\neg B \vee \neg B) \wedge (\neg B \vee A)$". This is equivalent to "$(\neg A \vee \neg B) \wedge (A \vee \neg B)$". Then, removing the disjunctions inside the brackets by applying de Morgan's rule, we get "$(\neg(A \wedge B)) \wedge \neg(\neg A \wedge B)$". This is the required form of the original sentence F.

[78]In the *Tractatus* the principle is effectively a criterion of sense – every meaningful proposition must be

as analogous to the difference between the positive and negative print of a photograph. This idea had figured already in Wittgenstein's *Notebooks 1914–1916* where he is on record saying: "What I mean to say is that we *only* then understand a proposition if we know *both* what would be the case if it was *false* and what if it was *true*." [Wittgenstein (1979) [1914–1916]] Hintikka takes this as a convincing case for maintaining that "Wittgenstein consistently insisted that the understanding of not-p is implicit in the understanding of p." [Hintikka (1996), 41] The claim that both kinds of understanding are pictorial must be taken with a grain of salt to be palatable. In order to make the picture idea more tangible one would have to give a concrete construction method for the complex presentations of pictures and make precise what is involved in such a process of construction. If the *pictorial* element in the correspondence between propositions and facts is to be taken seriously, and not as a mere allegoric way of speech, we are certainly justified to ask for a more detailed description of the constructive features of such a picturing. Hintikka does not provide many clues as to how to set up such constructions, although he makes an allusion to the work of Hao Wang where the pictorial dimension of logical description is explicated through *domino problems*. I will come to this issue shortly.

Still, there is one conceptual problem that needs to be clarified in order to make Wittgenstein's account of the picture theory of propositions coherent. The problem concerns the fact that according to Wittgenstein the syntactical form of a proposition of natural language does not reveal its logical form. The syntactical form is accommodated into the theory because

> [i]n everyday language it very frequently happens that the same word has different modes of signification — and so belongs to different symbols — or that two words that have different modes of signification are employed in propositions in what is superficially the same way. Thus the word 'is' figures as the copula, as a sing for identity, and as an expression for existence; 'exists' figures as an intransitive verb like 'go', and 'identical' as an adjective; we speak of *something*, but also of *something's* happening. (In the proposition, 'Green is green' — where the first word is the proper name of a person and the last an adjective — these words do not merely have different meanings: they are *different symbols*.) [3.323] . . .

To eliminate the ambiguities Wittgenstein mentions above it is necessary to construct "an ideal language", a language which is constructed in such a manner that its syntactical forms reflect its logical forms. This is needed in order to retain the picturing capacity of the language.

> In order to avoid such errors we must make use of a sign-language that excludes them by not using the same sign for different symbols and by not using in a superficially similar way signs that have different modes of signification:

either true or false.

> that is to say, a sign-language that is governed by *logical* grammar — by logical syntax. (The conceptual notation of Frege and Russell is such a language, though, it is true, it fails to exclude all mistakes.) [3.325]

Now there arises a problem. How can the propounded picture relation between a proposition and a state of affairs be preserved if the syntactical form of a proposition fails to reveal its logical form? Wittgenstein has an answer ready at hand:

> In order to recognize a symbol by its sign we must observe how it is used with a sense. [3.326]

Furthermore, towards the end of the *Tractatus* there is a proposition which elaborates on this:

> The *application* of logic decides what elementary propositions there are.
> What belongs to its application, logic cannot anticipate.
> It is clear that logic must not clash with its application.
> But logic has to be in contact with its application.
> Therefore logic and its application must not overlap. [5.557]

These quotations are all in accordance with the interpretation that the pictorial form is related with the syntactical form. However, the *use* of a given propositional expression is the required pragmatic ingredient that renders the pictorial relation complete and comprehensible. Without this pragmatic ingredient there is no picturing of a state of affairs. Therefore the "syntactical clothing" (Hintikka) of a sentence is seen to be insufficient to fulfill the task of extra-linguistic reference. As Wittgenstein crystallizes it:

> What signs fail to express, their application shows. What signs slur over, their application says clearly. [3.262]

Here the "signs" are to be comprehended as signs comprising the expression of a given proposition. The whole gist of the picture theory is that the picturing relation is defined between propositions and states of affairs. In this respect some commentators have went astray in thinking that the pictorial relation establishes a one-to-one correspondence between a given singular term in an expression and an object of the world. For instance, given a proposition aRb designating a relation between the two objects a and b, the terms 'a', 'R', and 'b' are taken to 'picture' *separately* the particular objects a, R, and b of the world.[79] Peter Carruthers, for example, has interpreted the picture theory of Wittgenstein along these lines and made on this basis a distinction between a stronger and a

[79] Some authors speak about "reference" in this connection which commits a double mishap in misunderstanding the nature of the picturing relation *and* advocating a semantical mode of speech that does not straightforwardly apply in the context of Wittgenstein's theory. That the talk about "the reference of an individual term" is problematic here is due to the fact that under Wittgenstein's account individual terms

weaker version of the doctrine of isomorphism. According to the stronger version not only names but also the significant relations between them will stand in (the picture) relation of reference to the world. In the weaker version the assumption of the significant relations having reference to the world is dropped. Carruthers himself advocates an interpretation of Wittgenstein's picture theory where a non-referential semantics for predicative expressions is embodied in the *Tractatus*. [Carruthers (1990)] But this is, I maintain, an erroneous line of interpretation. We cannot make sense of the picturing relation in this mitigated case neither.[80] Indeed, the reading of Hintikka makes it even comprehensible why the "rejection of the picture theory by Wittgenstein" is not a rejection at all. Wittgenstein only dropped the requirement of defining pictorial form in terms of the syntactical form of a proposition. Therefore, only the syntacticity thesis (6) was dropped, leaving the other theses (1–5) more or less intact. In this sense Wittgenstein never gave up the picture theory, as is made evident by the considerations appertaining to the relation between him and the members of the Vienna Circle which are taken up later in this chapter.

I will now revert to the case Hintikka has taken up in connection with his conception of Wittgensteinian pictures as "jigsaw puzzles". As I remarked earlier, it would certainly make the whole idea of picturing more coherent and comprehensible, if the "picture" aspect were taken seriously in it, and not only as an involved way of alluding to a somewhat mysterious relationship between linguistic expressions and states of affairs. Hintikka describes Wittgenstein's approach to this question as "jigsaw puzzle theory of the nature of logic". [Hintikka (1996), 48] What this means is that the sole foundation upon which logic is constructed is a collection of names, corresponding to a collection of simple objects. (Hintikka risks here sliding towards a simplistic referential interpretation of the picture theory.) In addition, each object is given together with a specification of the ways in which it can or cannot be combined with the other objects. Hence the term "jigsaw puzzle". Now, the next step is to ask how the expressive force of the familiar first-order logic could be built on such foundations. *Prima facie* such a task seems impossible. But it is *not* impossible: in connection with certain investigations of the relations of combinatorial games and logic, Hao Wang and his collaborators have shown that from a relatively simple setting of a particular type of a combinatorial game, a surprisingly rich

never have senses on their own. Or, as is said in the proposition [3.142]: "Only facts can express a sense, a set of names cannot." Therefore only a propositional sign (as a whole) has a sense, and can be involved in the picturing relation. Although the idea that reference or designation is a kind of 'picturing' seems *prima facie* plausible, this kind of a relation has very little to do with the picture relation that Wittgenstein envisaged. Consider the simple example "Bertrand Russell is talking with Alfred Ayer". What would a 'picture' of Bertrand Russell abstracted from the situation in which he is presented to us (as sitting comfortably in a chair discussing with Ayer) be like? For sure, the name "Bertrand Russell" designates the individual Bertrand Russell, but the picturing relation is not defined for individual constants. When speaking in the framework of his 'picture theory', Wittgenstein is not talking about reference of proper names in particular or about reference in the model-theoretical sense in general. He is talking about pictures as some kind of generic relations between propositions and states of affairs. Although the individual terms can be interpreted as designating certain concrete individuals, the logical framework Wittgenstein is working in does not admit of such an interpretation. If we stick to the Fregean idea that the sense of a term determines its reference, we do not, in the case of Wittgenstein's picture theory have a well-defined notion of reference (of individual terms). Only propositional signs enter into the picturing relation.

[80] The only coherent reading in my opinion being the one referred to in the preceding footnote.

structure can be derived. Moreover, some of these structures are seen to be equivalent to the structure (or a relevant substructure) of first-order languages. Especially interesting is the case where Wang has shown that the decision problem of the entire first-order logic can be expressed in the context of so-called domino-problems. The description of Wang's domino problem is as follows:

DOMINO PROBLEM involves assembling three coloured tiles called domino types to form a block that can be infinitely extended with colours matching on all adjacent edges (*Fig.*). It is assumed that the player has an infinite quantity of each domino type and that no domino can be rotated in two dimensions. The problem is solved by finding a rectangular block in which the colour sequence on the top edge is the same as that on the bottom edge and the sequence on the left edge is the same as that on the right. Such a unit can be repeated in all directions to fill an infinite plane. [Wang (1990), 196]

The correspondence between a domino problem and a decision problem is the following: for any given set of dominoes we can find a corresponding "AEA-formula"[81] such that the set has a solution if and only if the formula is not self-contradictory. In other words, a domino problem can be transformed into a problem of determining whether a corresponding AEA-formula is self-contradictory or not. From this it follows that since the general domino problem is unsolvable, there is no general method for deciding whether an arbitrary AEA-formula is self-contradictory.

The significance of this result is summarized by Wang himself as follows:

> The result is useful because the complexity of formulas in logic is to a large extent measured by the number and order of quantifiers, and the formulas of logic are often put into different classes according to the structure of the quantifiers [the quantifier prefix]. It is surprising that as simple a class as that of AEA-formulas (with three quantifiers only) is undecidable. In fact, with this result the decision problem for all quantifier classes are answered. Given any string of quantifiers we can now tell if the class of formulas determined by it is decidable.
> The decision problem of logic is significant because all mathematical theories can be formulated in the framework of elementary logic. The question of whether or not a formula (F) can be derived from a set of axioms (A) is reduced to deciding if the logical formula "A but not F" is not self-contradictory. In this sense all mathematics is reducible to logic. Indeed, one measure of the complexity of a mathematical problem is given by the structure of its corresponding formula in logic. It is therefore an important enterprise to determine the complexity of various classes of logical formulas. [*ibid.*, 208]

[81] A first-order formula having the quantifier prefix AEA, denoting in effect a formula of the type $\forall x \exists y \forall z P(x, y, z)$, where P is any complex predicate not involving quantifiers (with possibly a sequence $a_0, \ldots, a_n$ of individual constants, their number being dependent on the arity of P.

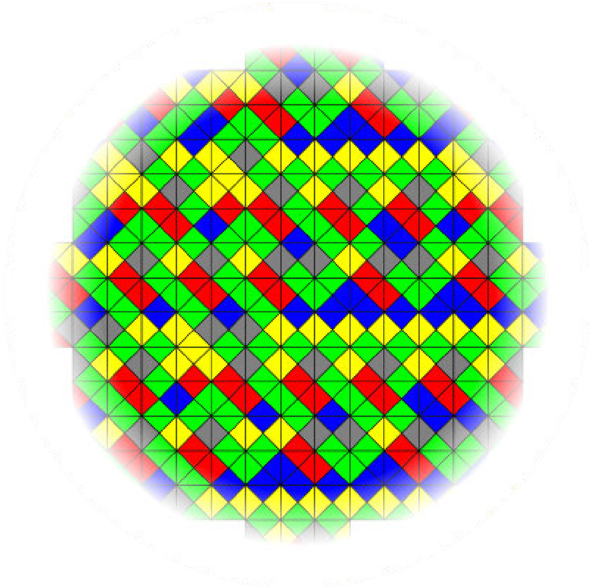


Figure 4.4: Wang Tiles. ©Wikipedia.

Hintikka relates his own theory of the distributive normal forms to Wang's account of formulas in first-order logic to illustrate the coherence and fruitfulness of the idea of "jigsaw puzzles". Indeed, given the ingredients of Hintikka's DNFs,

$$\begin{aligned} &\bigwedge_{i\in I}(\exists x)C_i[x] \quad \& \\ &(\forall x)\bigvee_{i\in I} C_i[x] \quad \& \quad \bigwedge_{j}(\pm)A_j \end{aligned} \tag{4.1}$$

a case can be made for the resemblance of the accounts of Wang and his in this particular respect. The last term on the second line is a conjunction of atomic formulas (either negated or unnegated). The interpretation of the rest of the terms is that they effectively list all the different kinds of individuals there are. Moreover, each $C_i[x]$ has the identical form:

$$\begin{aligned} &\bigwedge_{j\in J}(\exists y)C_j[y,x] \quad \& \\ &(\forall y)\bigvee_{j\in J} C_j[y,x] \quad \& \quad \bigwedge_{k}(\pm)A_k[x] \end{aligned} \tag{4.2}$$

The consistency requirements for (4.2) are that (1) each x in (4.1) "has to find a slot among the $y's$ in each conjunction occurring in (4.2)." As Hintikka remarks, "this is very much like saying that any two jigsaw puzzle pieces must be compatible, capable of being fitted into one and the same piece. Moreover, for each y in (4.2) there must be a compatible x in (4.1). This is very much like saying that each gap left by any given jigsaw puzzle piece must be capable of being filled by one of the available pieces." [Hintikka (1996), 50]

This is, then, the more concrete sense of the picturing relation that we have been looking for. Of the two interpretations of Wittgenstein's picture theory (Wang's domino-scheme and Hintikka's distributive normal forms), the demonstration of the consistency of a given first-order formula by means of a domino game seems more attractive and illustrative. As we are going to see, the picture theory underlies much of the philosophical work of Vienna Circle. This is especially so with the logical work of Carnap. It comprises an overarching presupposition of nearly all of Carnap's constructional enterprises in logic and language engineering tasks far in to the 1930s when Carnap adopts a more liberal and pragmatic attitude towards concept formation and theory construction. The grasp of the picture theory over Carnap's thinking is particularly strong in his *Der Logische Aufbau der Welt*. This book is also one of the clearest expositions of the program of rational reconstruction that Carnap had envisaged already in 1922. As this single book accommodates such a prominent position in the development of Carnap's thought and in the subsequent discussions among the members of the Vienna Circle as well as within contemporary epistemology in general, it is necessary to turn to adduce its main lines of thought in more detail.

4.5.2 The objective of the *Aufbau*

> For my philosophical work the period in Vienna was one of the most stimulating, enjoyable, and fruitful periods of my life. My interests and my basic philosophical views were more in accord with those of the circle than with any group I ever found. From the very beginning, when in 1925 I explained in the Circle the general plan and method of *Der Logische Aufbau*, I found a lively interest. When I returned to Vienna in 1926, the typescript of the first version of the book was read by members of the Circle, and many of its problems were thoroughly discussed. [Carnap (1963), 20]

The point of departure in the *Aufbau* is the epistemological problem alluded to above in connection with Russell. It is raised by the observation that knowledge, for any given agent, begins in the stream of personal experience of that individual. Now, individual experience is inherently something that we take as representing the paradigm of *subjective*: the qualitative contents of individual experience are wholly private to the person whose experience it is. This means that the peculiarities of the individual experience cannot be communicated to others. Thus, we are led to the essential problem of the *Aufbau*: how, then, "is it possible [...] to attain an intersubjective, *objective world* that is conceptually comprehensible and, indeed, as one identical for all subjects"? [Carnap (1967) [1928], §2] Carnap's solution draws on two essentially neo-Kantian ideas (although the details of the solution depend equally on Husserl's conception of knowledge acquisition)[82]: (1) there are structural features in common among the individual streams of experience. (2) One can define the objects of knowledge wholly on the basis of such shared structural features of experience. [Richardson (1998), 32]

Thus, underlying Carnap's solution of this problem is a conspicuously structuralist view of objectivity, clearly influenced by the developments in the foundations of geometry that

[82]See especially [Haddock (2008)].

constituted the object of his very first philosophical investigations. Carnap gives a lucid description of his view in §66 of the *Aufbau*:

> Our problem now is how science can arrive at intersubjectively valid assertions if all its objects are to be constructed from the standpoint of the individual subject, that is, if in the final analysis all statements of science have as their object only relations between "my" experiences? Since the stream of experience is different for each person, how can there be even one statement of science which is objective in this sense (i.e., which holds for every individual, even though he starts from his own individual stream of experience)? The solution to this problem lies in the fact that, even though the *material* of the individual streams of experience is completely different, or rather, altogether incomparable, since a comparison of two sensations or two feelings of different subjects, as far as their immediately given qualities are concerned, is absurd, certain *structural properties* are analogous for all streams of experience. Now, if science is to be objective, then it must restrict itself to statements about such structural properties, and, as we have seen earlier, it can restrict itself to statements about structures, since all objects of knowledge are not content, but form, and since they can be represented as structural entities. [Carnap (1967), §66]

In paragraph §119 Carnap gives the two basic theorems of construction theory that express the overall purpose of the entire program of reconstruction. The first of these states, on the one hand, that *each scientific concept is either a class or a relation extension, which can be expressed through the basic relation(s) alone.* The second theorem states, on the other hand, that *each scientific statement is, in the final analysis, a statement about the basic relation(s).* [Carnap (1961 [1928]), 187] To give an example of the complexity that a constructional definition yields even in connection with a relatively simple statement, consider the Theorem 6 of Carnap's book, essentially stating that *the color solid is three-dimensional*. Carnap proceeds to adduce the example in the following terms:

> Let us clarify this thesis [the second basic theorem] with the example of Th. 6 of the three dimensionality of the color solid. With the aid of the constructional definition of *Proxcol*, Th. 6 can be transformed, through substitution, into the sentence:
>
> $$3 \quad \text{Dnhomvic} \quad (\breve{\epsilon} \mid \text{Sim} \mid \epsilon) \quad \triangleright \quad \text{Color}\,[\dots]$$
>
> Through step-by-step substitutions on the basis of the definitions of color, *Colid*, *Colidprox*, *Proxpl*, *place*, *Excl*, *sight*, *sense*, *Sim*, *qual*, *similcirc*, *Ps*, and a formal simplification we finally obtain from [the expression above] the following form for Th. 6; in this form, "*Rs*" is the only nonlogical symbol (*Q*, *x* and the Greek letters are variables; the other symbols are logical constants):

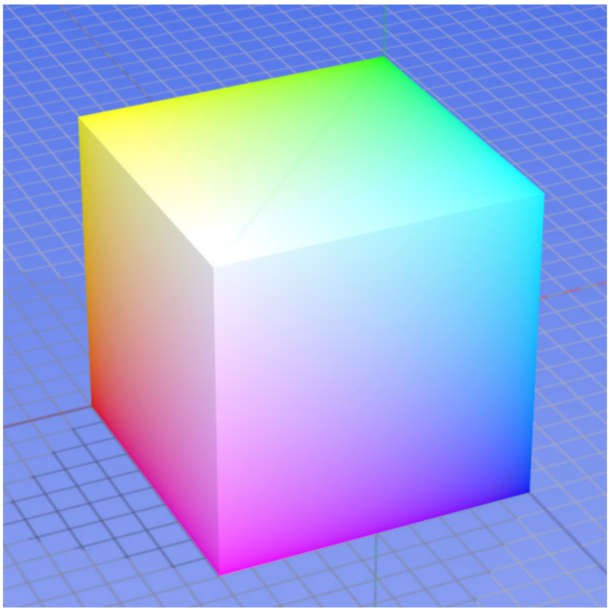


Figure 4.5: The Color Solid Cube. ©Wikipedia.

$$(\exists \mathrm{Q}, \nu).3\mathrm{Dnhomvic}(\breve{\epsilon} \mid \mathrm{Q} \mid \epsilon) \rhd \mathrm{Abstr}'\{\hat{\alpha}\hat{\beta}((\exists \chi, \lambda, \mu).\chi\breve{\epsilon} \mid \mathrm{Q} \mid \epsilon\lambda.\chi\breve{\epsilon} \mid \mathrm{Q} \mid \epsilon\mu.$$

$$\lambda\breve{\epsilon} \mid \mathrm{Q} \mid \epsilon\mu.\chi, \lambda, \mu\epsilon\hat{\xi}\{\exists!\xi : (\exists\rho).\rho\epsilon\nu.\xi = \rho - s'(\nu - [\rho])\}.\alpha\epsilon\chi.\beta\epsilon\lambda.\mu \bigcap \vec{\mathrm{Q}}'\alpha = \mu \bigcap \vec{\mathrm{Q}}'\beta)\}_{po}.\nu =$$

$$\mathrm{Simil}'((\mathrm{E} \times \bigcup I) \rhd \hat{\alpha}\{(\exists\mu).\mu\epsilon\mathrm{Abstr}'\mathrm{Q}_{po}.\mathrm{Dnp}(5, \mu, \alpha, \mathrm{Vicin}'\mathrm{Q})\}).\mathrm{Q} = \hat{\alpha}\hat{\beta}(\alpha, \beta\epsilon\hat{\zeta}\{(\gamma) :$$

$$\gamma\epsilon\mathrm{Simil}'(\mathrm{Rs} \bigcup \check{\mathrm{R}}\mathrm{s} \bigcup \mathrm{Rs}^0).\mathrm{Nc}'(\zeta \bigcap \gamma)/\mathrm{Nc}'\zeta > \frac{1}{2}. \supset .\zeta \subset \gamma : .(\times) : \times \sim$$

$$\epsilon\zeta. \supset .(\exists\delta).\delta\epsilon\mathrm{Simil}'(\mathrm{Rs} \bigcup \check{\mathrm{R}}\mathrm{s} \bigcup \mathrm{Rs}^0).\alpha \subset \delta.\times \sim \epsilon\delta\}.\alpha \uparrow \beta \subset \mathrm{Rs} \bigcup \check{\mathrm{R}}\mathrm{s} \bigcup \mathrm{Rs}^0)[\ldots]$$

[*ibid.*, §187]

The upshot of Carnap's example is that the empirical statements concerning the three-dimensional color solid can be formulated, in the given basis, as statements about "*a certain, purely formal, though very complicated, property of the basic relation Rs.*" [*ibid.*] Moreover, the singular case of the color solid can be generalized to yield the principle that underlies the constructional program: "In the same way, *all empirical statements of science can be expressed as statements about purely formal properties of the basic relation(s).* This holds generally, no matter which basic relations and no matter what constructional system may be chosen."[83] [*ibid.*]

A central aspect of the construction theory that appertains to the characterization of visual sensations was the assumption of the two-dimensionality of visual space. Before *Aufbau* this had more or less been taken as given by Carnap, an intuition too evident to

[83]Emphasis in the original.

merit closer scrutiny. But this was a mistake, as Carnap came to admit later. A more critical attitude towards this assumption is shown in the *Aufbau* where the two-dimensional visual field is constructed (in the §89 and §117) from the primitive relations.

A problem that Carnap proceeds to tackle in §126, viz. that of assigning colours to world points, is different from the one that he had tackled in 1922. Whereas the earlier account had proceeded from an unorganized, chaotic impressions of the two-dimensional "primary world" to the three-dimensional "secondary world", the account given in the *Aufbau* was not a description of an 'ontological transition' of this sort. The question could now be seen as entirely mathematical, consisting essentially of the formal mapping of a two-dimensional manifold to a three-dimensional one. Here, Carus has suggested, with circumstantial evidence provided by Michael Friedman, that Carnap may have been influenced in this respect by Karl Menger, the mathematician that organized the parallel discussion group to the Schlick Circle, the "Mathematics Colloquium". Menger had presented his researches on dimension theory in a monograph *Dimensionstheorie* published in 1928. Friedman regards it as probable that Carnap incorporated these ideas in the *Aufbau* without explicitly mentioning it. The circumstantial evidence I alluded to is a mention of Menger's monograph in *Abriss der Logistik* [Carnap (1929), 77] which has been located by Friedman. As it happens, this conjecture is made highly plausible by another, much more direct source, viz., the statement of Menger himself in his *Reminiscences of the Vienna Circle and the Mathematical Colloquium*. [Menger (1994)] There is a passage in which Menger relates that

> Rudolf Carnap had visited Vienna in the spring of 1925 before I left the city for Amsterdam. At that time, he and I had a long talk about his work and about the theories of dimension and of curves which I was then developing and in which he was greatly interested since he had just finished a booklet about space.[84] [Menger (1994), 63]

This passage renders it indisputable that Carnap was in fact influenced by Menger's ideas on dimension theory already in 1925 (instead of 1929). Thus, it becomes highly likely that the assignment of colours to world points in §126 of *Aufbau* is actually influenced by Menger although Carnap does not mention him.

4.5.3 The crisis of rational reconstruction

The central contention about the conception of language in the Vienna Circle at the time was that although the basic approach propounded in Wittgenstein's *Tractatus* was indispensable (including among others, the idea of non-factuality of analytic sentences), it

[84] Menger is here referring to *Der Raum* (according to [Menger (1994)]). However, a lapse of three years would make it hard to justify talking about Carnap having "just finished a booklet about space". I think it is more probable that the 'booklet' Menger is alluding to (evidently on the basis of Carnap's own reports) is the article "Dreidimensionalität des Raumes und Kausalität: Eine Untersuchung über den logischen Zusammenhang zweier Fiktionen" [Carnap (1924)]. At least this would suit more exactly the time and the subject of their discussion.

would have to be expanded to accommodate also physical laws and 'elucidations' concerning metalinguistic facts. Carnap naturally shared this contention. But his relationship to Wittgenstein was founded on a 'bipolarity' that comprised two contrary conceptions of the importance of Wittgenstein's view. On the one hand, Carnap fundamentally agreed with Wittgenstein's conception of analytic truth. In Carnap's view it was Wittgenstein who, more than any other, had made clear the nature of logical truth itself: indeed, the essential insight of Wittgenstein was that the truths of logic are tautologies that hold in all possible worlds (or, in accordance with Wittgenstein's original intentions, in all possible states of affairs) and hence are inherently without content, saying nothing about the world. This idea figured prominently in Carnap's thinking:

> For me personally, Wittgenstein was perhaps the philosopher who, besides Russell and Frege, had the greatest influence on my thinking. The most important insight I gained from his work was the conception that the truth of logical statements is based only on their logical structure and the meaning of their terms. Logical statements are true under all conceivable circumstances; thus their truth is independent of the contingent facts of the world. On the other hand, it follows that these statements do not say anything about the world and thus have no factual content. [Carnap (1963), 25]

It is evident that the central point of agreement between Carnap's thinking and the ideas presented in the *Tractatus* is precisely the conception of the tautologous nature of logical truths (they diverged with respect to mathematical truths).[85] On the other hand, there was also a critical point of disagreement: whereas Wittgenstein had built his treatment in the *Tractatus* around the contrast between the sayable and unsayable, leading his exposition towards the well-known *coda* of "Wovon man nicht sprechen kann, darüber muß man schweigen", Carnap vehemently advocated the view that the analysis of the structure of language, especially the linguistic forms in which we express and state facts, are

[85]Carnap illustrates this agreement by describing the essential contribution of Wittgenstein to the articulation of analyticity as follows: "The conception of the nature of mathematics which we developed in the discussions of the Vienna Circle came chiefly from the following sources. I had learned from Frege that all mathematical concepts can be defined on the basis of the concepts of logic and that the theorems of mathematics can be deduced from the principles of logic. Thus the truths of mathematics are analytic in the general sense of truth based on logic alone. The mathematician Hans Hahn, one of the leading members of the Circle, had accepted the same conception under the influence of Whitehead['s] and Russell's work, *Principia Mathematica*. Furthermore, Schlick, in his book *Allgemeine Erkenntnislehre* (1918), had clarified and emphasized the view that logical deduction cannot lead to new knowledge but only to an explication or transformation of the knowledge contained in the premises. Wittgenstein formulated this view in the more radical form that all logical truths are tautological, that is, that they hold necessarily in every possible case, therefore do not exclude any case, and do not say anything about the facts of the world. Wittgenstein demonstrated this thesis for molecular sentences (i.e., those without variables) and for those with individual variables. It was not clear whether he thought that the logically valid sentences with variables of higher levels, e.g., with variables for classes, for classes of classes, etc., have the same tautological character. At any rate, he did not count the theorems of arithmetic, algebra, etc., among the tautologies. But to members of the Circle there did not seem to be a fundamental difference between elementary logic and higher [order] logic, including mathematics. Thus we arrived at the conception that all valid statements of mathematics are analytic in the specific sense that they hold in all possible cases and therefore do not have any factual content." [Carnap (1963), 46]

the most important objects of the "logical syntax" which he called theoretical science in the strictest sense:

> Furthermore, there is a divergence on a more specific point which, however, was of great importance for our way of thinking in the Circle. We read in Wittgenstein's book that certain things show themselves but cannot be said; for example the logical structure of sentences and the relation between the language and the world. In opposition to this view, first tentatively, then more and more clearly, our conception developed that it is possible to talk meaningfully about language and about the relation between a sentence and the fact described. [...] [I] pointed out that only the structural pattern, not the physical properties of the ink marks, were relevant for the function of language. Thus it is possible to construct a theory about language, namely the geometry of the written pattern. This idea led later to the theory which I called "logical syntax" of language. [Carnap (1963), 29]

This, then, is the background against which the aspiration towards an expanded conception of language must be understood. The conception that Carnap had propounded in the *Aufbau*, namely that all concepts were to be constituted by explicit definitions,[86] was beginning to be undermined. Carnap had come to realize that even some fundamental concepts of the 'pure' sciences, such as those of arithmetic and set theory could only be attained axiomatically. The challenge that resulted from this insight was that it now became a problem to develop a language that would at the same time fulfill the requirements of implicit definability of concepts (i.e. axiomatic representability of a system of concepts) and the clarity and non-ambiguity of the concepts (Carnap uses the term "*eindeutig*"). It was actually Moritz Schlick who instigated Carnap to tackle this problem, as he had already in the *Allgemeine Erkenntnislehre* (1918) addressed the problem of the seemingly unbridgeable gap between axiomatically definable concepts and observable entities.[87] This was the point of departure for Carnap's 1927 paper in which the contrast between "proper [*eigentliche*]" (explicitly definable) concepts and "improper [*uneigentliche*]" (implicitly definable) concepts is for the first time presented:

---

[86] "Unlike other conceptual systems, a constructional system undertakes more than the division of concepts into various kinds and the investigation of the differences and mutual relations between these kinds. In addition, it attempts a step-by-step derivation or 'construction' of all concepts from certain fundamental concepts, so that a genealogy of concepts results in which each one has its definite place. It is the main thesis of construction theory that all concepts can in this way be derived from a few fundamental concepts, and it is in this respect that it differs from most other ontologies." [Carnap (1928 [1967]), §1] Cf. also [*ibid.*, §2].

[87] "[...] kurz, durch die konkrete Definition wird der Zusammenhang der Begriffe mit der Wirklichkeit hergestellt, sie zeigt in der anschaulichen oder erlebten Wirklichkeit dasjenige auf, was nun durch den Begriff bezeichnet werden soll. Die implizite Definition dagegen steht nirgends in Gemeinschaft oder Verbindung mit der Wirklichkeit, sie lehnt sie absichtlich und prinzipiell ab, sie verharrt im Reich der Begriffe. Ein mit hilfe impliziter Definition geschaffenes Gefüge von Wahrheiten ruht nirgends auf dem Grunde der Wirklichkeit, sondern schwebt gleichsam frei, wie das Sonnensystem die Gewähr seiner Stabilität in sich selber tragend. Keiner der darin auftretenden Begriffe bezeichnet in der Theorie ein Wirkliches, sondern sie bezeichnen sich gegenseitig in der Weise, daß die Bedueutung des einen Begriffes in einer bestimmten Konstellation einer Anzahl der übrigen besteht." [Schlick (1918), 35]

> Logically, the implicitly defined concepts differ so radically from proper concepts that one may well have doubts about even calling them 'concepts' at all. We will retain this name, however, in view of common usage, especially within mathematics ... [where] one talks as if one were dealing with concepts — 'point', 'line', 'between', etc. — that meet all the requirements of a legitimate concept. Since this is not the case, we will limit our terminological concession to usage by calling implicitly defined concepts '*improper concepts*'.[88] [Carnap (1927), 366-367]

There are some conspicuous similarities between the approach Carnap is taking here and the view of Frege. Frege, as is well known, had established that a set of axioms that was used to implicitly define a concept, comprised a basis for an explicit definition of a *second-order* concept. This concept, being of of course of a higher type, could not be substituted for the implicitly definable original concept. Moreover, as Carnap naturally recognized, whereas explicitly defined concepts were in a definite sense *constant* terms (under a linguistic representation), the implicitly defined concepts were *variables*, admitting of several possible interpretations (ranging over different models), and could thus be of use in the empirical sciences. [*ibid.*] This gave arise to the following problem, however:

> Empirical concepts are constituted step-by-step in the systematic construction [*Aufbau*] of our world-knowledge. Each empirical concept, as a component of this structure, has a direct connection to reality. In contrast, the improper concepts hang in the air, so to speak, awaiting instructions. They are introduced by an axiom system, but that system does not relate directly to anything real. The axioms of this system and theorems deduced in it do not properly form a theory (as they are not actually *about* anything in particular), but rather just a theory-schema, an empty framework for possible theories.[89] [Carnap (1927), 372]

As Carus remarks, there are actually two partial problems relating to this question of *Eindeutigkeit* (disambiguation): (1) the question of determining whether a given particular object falls under a given concept, and (2) the question about the criteria for establishing whether a concept picks out a unique collection of objects.[90] [*ibid.*] With regard to the first question, Carnap remarks:

[88]Translated in [Carus (2007), 192]

[89]Translated in [*ibid.*].

[90]These difficulties reflect the ambiguity in the philosophers' received notion of "implicit definability". This is the view according to which philosophers have sometimes referred to definitions by postulates as *implicit definitions*. A notable examples of this is e.g. [Braithwaite (1953)]. But, as Veikko Rantala has remarked in connection with his inquiries on definability, "it is not clear to what extent it is legitimate to consider such a procedure as yielding a genuine definition of the notions involved [...], and in any case definitions by postulates [axioms] must be distinguished sharply from the logical notion of 'implicit definability' [...]." [Rantala (1977), 11] (N.B. The semantic counterpart to the syntactical notion of *implicit definability* is provided by Beth's theorem which states that the following are equivalent: (i) For every model $\mathcal{M}$ for $\mathcal{L}$, $|\{X \subseteq M^m | (\mathcal{M}, X) \models T(P)\}| \leq 1$; (ii) There is a formula $\varphi(\bar{x})$ of $\mathcal{L}$ such that $T(P) \vdash \forall\bar{x}(P\bar{x} \iff \varphi(\bar{x}))$. (i) expresses the formalization of the notion of *implicit definability*.)

> It belongs to the essence of a proper concept that for every object we can in principle decide whether it falls under that concept or not; and for sufficiently well known objects the decision can be carried out in practice as well. For the empirical concept horse, for instance, and any visible object we can unambiguously [*eindeutig*] decide — insofar as the concept has sufficiently sharp boundaries and the object is sufficiently well known — whether the object satisfies the concept, i.e., whether or not it is a horse. But for an improper concept the question whether a particular object falls under it is not decidable and thus has no sense.[91] [Carnap (1927), 367]

Carnap devoted the years 1927–1930 primarily for writing a comprehensive treatise focusing on the second question. Having settled in Vienna and having grasped and to a large part assimilated the view of Wittgenstein's *Tractatus*, it now became a priority for him to address in a more concentrated and sustained manner the issue of definability. The treatise was tentatively titled *Untersuchungen zur allgemeinen Axiomatik*[92] and its goal was to construct a system in which implicit definitions could be converted to explicit ones. The requirement was essentially to set up all axiomatic systems within a 'foundational discipline' [*Grunddisziplin*] which would comprise 'absolute' or explicitly defined concepts. Carnap sometimes called his project in this book by the term "Metalogik", although it was not in a strict sense a treatment of metalogic proper.[93] The central tenets of the *Untersuchungen* were clearly stated by Carnap:

> Durch die neueren Untersuchungen über allgemeine Eigenschaften von Axiomensystemen, wie: Vollständigkeit, Monomorphie (Kategorizität), Enstscheidungsdefinitheit, Widerspruchsfreiheit u.a., und über die Probleme der Kriterien und der gegenseitigen Beziehungen dieser Eigenschaften ist immer deutlich geworden, daß die Hauptschwierigkeit der Probleme in der ungenügenden Schärfe der verwendten Begriffe liegt. Das wichtigste Erfordernis für eine fruchtbare Behandlung der vorliegenden Probleme ist einerseits eine ausdrückliche *Festlegung der* jeweils verwendeten *logischen Basis*, die meist nicht genau umrissen wird, und andererseits eine *Aufstellung scharfer Begriffsbestimmungen* aufgrund dieser Basis. Es soll hier besucht werden, diese beiden Forderungen zu erfüllen und dann durch die Ableitung einer Reihe von *Lehrsätzen der allgemeinen Axiomatik* die Fruchtbarkeit des gelegten

[91]Translated in [*ibid.*, 193].

[92]Originally comprising two parts, the first of which has quite recently been edited by Thomas Bonk and Jesus Mosterin, and published in 2000. [Carnap (2000)]

[93]Indeed, the course which he held at the University of Vienna during the writing process and which drew on the ideas of the manuscript to be published was actually called "Philosophical Foundations of Arithmetic". Hao Wang has commented on the significance of these lectures for the development of the ideas of Gödel: "Carnap offered a course, 'The Philosophical Foundations of Arithmetic', in the winter semester 1928/1929 (listed in the catalogue of the University of Vienna, two hours weekly), which was undoubtedly the one on 'metalogic' attended by G. It was probably only at the beginning of 1929 that the first edition of Hilbert-Ackermann (published in 1928) [*Grundzüge der theoretischen Logik*] became available to G and Carnap. Both the course and the book included questions of a mathematical character that bear directly on basic logical concepts." [Wang (1987), 22]

> Fundamentes zu erweisen. Unter "allgemeiner Axiomatik" ist dabei die Theorie der allgemeinen, logisch-formalen Eigenschaften von Axiomensystemen und der Beziehungen zwischen Axiomensystemen verstanden, zur Unterscheidung etwa von einer "speziellen Axiomatik", die sich mit bestimmten einzelnen Axiomensystem befaßt (z. B. einem Axiomensystem der euklidischen Geometrie, der Mengenlehre, usw.).[94] [Carnap (2000) [1927–1930], 59]

Carnap then turns to state more explicitly the underlying idea of his "general axiomatics". Notice the demarcation between an interpreted set of *logical* symbols and the non-interpreted set of *descriptive* symbols of a system.

> Aus einen vorgelegten Axiomensystem können nur Folgerungen gezogen werden, wenn noch allgemeine Regeln des Folgerns gegeben werden. Jede Behandlung und Prüfung eines Axiomensystems setzt also eine Logik voraus, und zwar eine inhaltliche Logik, d. h., ein System von Sätzen, die nicht Bloße Zeichenzusammenstellungen sind, sondern eine Bestimmte Bedeutung haben. Denn sonst würden sie uns nicht in den Stand setzen, zu handeln; und Deduzieren ist Handeln, denn es bedeutet: aus vorgegebenen Zeichenzusammenstellungen nach festen Regeln andere Zusammenstellungen bilden.
>
> Im unterschied zu den logischen Zeichen haben die "*Grundzeichen*", die Zeichen der "*Grundbegriffe*" eines Axiomensystems keine bestimmte Bedeutung. Denn das ist ja gerade das Wesentliche eines Axiomensystems, daß es nicht auf ein bestimmtes Anwendungsgebiet festgelegt ist, daß es nicht von Gegenständen handelt, die an sich schon bestimmt sind, sondern von Unbestimmtem, das seine einzige Bestimmung erst durch das Axiomensystem erhält. Hieraus geht hervor, daß das System der logischen Sätze, dessen Voranstellung wir für jede Axiomatik gefordert haben, nicht selbst ein Axiomensystem in dem hier gemeinten Sinne sein kann.[95] [*ibid.*, 60]

How could this conception of axiomatics be used within the empirical sciences? To give an example, an axiom system of the kind that Carnap presented in the *Abriss der Logistik*, for example, or the of the kind he envisaged even earlier in the beginning of the 1920s, could be interpreted straightforwardly by substituting explicit definitions for each implicitly defined basic concept in the system. In such a system of 'empirical' concepts

[94] Italics in the original.

[95] "Consequences can be drawn from a specified AS [axiom system] only if general rules of inference are given as well. So every treatment and appraisal of an AS assumes a logic, and indeed a contentful logic, i.e., a system of sentences that are not just combinations of signs, but have a particular meaning. For otherwise it wouldn't put us in a position to act; and deduction is action, for it means: constructing collocations of signs by fixed rules from other collocations of signs. In contrast to logical signs, the '*basic signs*', the signs of the '*basic concepts*' of an AS, have no definite meaning. For that is just the essential character of an AS — that it is not tied down to a particular area of application, that it deals not with objects determinate in themselves but with something indeterminate that gets its only determination through the axiom system. From this it emerges that the system of logical sentences we required to be in place prior to everything axiomatic cannot itself be an AS in the sense intended here." Translation by Carus [*ibid.*, 193–194].

[*Realbegriffe*], each implicitly defined concept should be empirically constituted in such a manner that the empirically constituted concept has the same formal character [*formale Beschaffenheit*] as the original improper concept given in the AS in question.[96]

The first aim, then, is the construction of objects; it is followed by a second aim, namely, the investigation of the nonconstructional properties and relations of the objects. The first aim is reached through convention; the second, however, through experience." [Carnap (1928 [1967]), §179] This results, then, in what Carnap calls a "genuine theory":

> Through the contact between the empirical concept and the axioms (the former satisfying the latter), a connection is created, by a single stroke, with the whole of the theory-schema resting on the axiom system. The blood of empirical reality streams in through the point of connection and flows into the most ramified capillaries of the hitherto empty schema, which is thereby transformed into a genuine theory.[97] [Carnap (1927), 373]

If the problem seemed to be 'straightforwardly solvable' in connection with 'empirical' concepts [*Realbegriffe*], then the case of mathematical concepts (the ones being implicitly definable in an axiomatic system) proved more recalcitrant. As was clear to Carnap, in order to succeed in his attempt to legitimate the applicability of mathematical concepts to the empirical reality, the framework of logical concepts that coöordinate the relations between logical and empirical concepts would have to be explicitly specifiable. Moreover, Carnap also saw a direct connection between the questions about completeness and the notion of mathematical truth, which connected his interest in the 'general axiomatics' with his interest in philosophy of mathematics more generally. Thus, contemporaneously with his work on the constructibility of empirical concepts in the *Aufbau*, Carnap was also deeply involved with profound work in logic and metamathematics. At the time, the questions associated with these areas were deep and perplexing. However, Carnap had a novel strategy at hand to tackle even these problems. This strategy grew out of Fraenkel's and subsequent work on the issue of the completeness of axiomatic systems. Whereas the research done by Hilbert and his school in the 1920s had mainly been concentrated on the issue of consistency of deductive systems (wherein the goal of obtaining absolute consistency proofs by syntactical means was most prominent), the question about completeness was addressed more fully and explicitly by Abraham Fraenkel. The first detailed discussion on completeness appears in the second edition of Frankel's *Einleitung in die Mengenlehre* (1923) in a long section titled "The Axiomatic Method". This discussion had a profound influence on Carnap, and as a consequence the two were shortly exchanging ideas on the topic, both in correspondence and in person. Indeed, in the article alluded to above, "Eigentliche und uneigentliche Begriffe" (1927), Carnap presents ideas related

[96]From a logical point of view, however, statements which are made about an object become statements in the strictest scientific sense only after the object has been constructed, beginning from the basic objects. For, only the construction formula of the object — as a rule of translation of statements about it into statements about the basic objects, namely, about relations between elementary experiences — gives a verifiable meaning to such statements, for verification means testing on the basis of experiences.

[97]Translated in [*ibid.*, 194]

to the exchanges with Fraenkel. Frankel did the same in the third edition of his *Einleitung* (1928). What provided the basis for the intensive exchanges between these two men? It was first and foremost the following conviction that they shared, i.e.: that *three* different notions of completeness needed to be distinguished. Fraenkel formulated these notions as follows:

> [T]he completeness of a system of axioms demands that the axioms encompass and govern the entire theory based on them in such a way that every question that belongs to and can be formulated in terms of the basic notions of the theory can be answered, one way or the other, in terms of deductive inferences from the axioms. Having this property would mean that one couldn't add any new axioms to the given system (without adding to the basic notions) so that the system was "complete" in that sense; since every relevant proposition that was not in contradiction with the system of axioms would already be a consequence and, thus, not independent, i.e., not an "axiom" ...
>
> Closely related to this first sense of completeness, but by far not as far reaching and easier to assess, is the following idea: ... In general, a number of propositions that are inconsistent with each other and that can, thus, not be provable consequences of the same system of axioms can nevertheless be compatible with that system individually. Such a system of axioms leaves open whether certain relevant questions are to be answered positively or negatively; and it does so not just in the sense of deducibility by current or future mathematical means, but in an absolute sense (representable by independent proofs). A system of axioms of that kind is, then, with good reason, to be called incomplete ...
>
> Quite different, finally, is another sense of completeness, probably characterized for the first time by Veblen ... According to it a system of axioms is to be called complete — also "categorical" (Veblen) or "monomorphic" (Feigl–Carnap) — if it determines the mathematical objects falling under it uniquely in the formal sense, i.e., such that between any two realizations one can always effect a transition by means of a $1-1$ and isomorphic correlation. [98] [Fraenkel (1928), 347–349]

These notions can be characterized more briefly and in an updated terminology as follows [*ibid.* 184–185]:

(1) A system of axioms S is *deductively complete* if and only if for every proposition P in the relevant language either P or not-P deducible from S.

(2) A system of axioms S is *semantically complete* if and only if there is no proposition P in the relevant language such that both S together with P [99] and S together with not-P [100] are satisfiable, i.e. have a model.

[98] Translated in [Reck (2007)].

[99] I.e., $S \bigcup \{P\}$.

[100] I.e., $S \bigcup \{\neg P\}$.

(3) A system of axioms S is *categorical* (or *monomorphic*, as opposed to *polymorphic*) if and only if all models of S are isomorphic.

The *Eindeutigkeit* of just such concepts was intended to be secured by a theorem that formed the centerpiece of the first part of the *Untersuchungen*. Stated briefly, the theorem (called the *Gabelbarkeitssatz* by Carnap) says that the concepts "non-categorical" ("*polymorph*") and "*gabelbar*" are equivalent. It implies a theorem that is more transparent: an axiom system is complete (*entscheidungsdefinit*) if and only if it is categorical (*monomorph*).[101] The underlying thought behind the introduction of this theorem is that if we can show that a given axiom system is categorical, then it is known to be complete as well, and hence any sentence containing the predicator standing for an implicitly defined concept is decidable, i.e. the law of excluded middle holds for all such sentences. To fully grasp the significance of the theorem it is necessary to cover some conceptual background and provide exact definitions for the relevant concepts as Carnap's definitions of them are not entirely congruent with modern usage.[102] Carnap's statement of the theorem is based on the following distinction between different senses of "completeness [*vollständigkeit*]": (i) "*monomorph*", (ii) "*nichtgabelbar*" and (iii) "*entscheidungsdefinit*". Carnap goes on to give a preliminary description of these notions:

> Die Möglichkeiten scharfer Definitionen werden nachher ausführlich erörtert; vorläufig mögen die folgenden Andeutungen genügen, um verstehen zu lassen, was gemein ist: ein Axiomensystem heißt "monomorph", wenn nur eine Struktur zu ihm gehört [...]; ein Axiomensysteme heißt "nichtgabelbar", wenn eine Gabelung [...] nicht möglich ist; ein Axiomensysteme heißt "entscheidungsdefinit", wenn für jede formale Aussagenfunktion mit denselben Variablen gilt, daß entweder siel selbst oder ihr Negat eine Folgerung des Axiomensystems ist.[103] [Carnap (2000), 127–128]

Carnap's exposition uses a notational device that is confusing if its basic idea is not explicitly spelled out. The essential insight of Carnap is that the primitive symbols of an axiom system may be treated as variables, and that by conjoining the axioms (by the operation

[101] Carus erroneously terms the latter theorem *Gabelbarkeitssatz*. [Carus (2007), 195] In [Awodey & Carus (1998)], however, the reference is correct. The theorem referred to here is explicitly named in [Carnap (2000), 137].

[102] Indeed, it might seem that the theorem is plainly false, since, e.g., second-order Peano arithmetic fails, though categorical, to be complete. [Carus (2007), 195n]

[103] Awodey and Carus give the following elucidations of these notions: "(1) *Decidable.* An axiom system f is *consistent* if for no propositional function g, both g and $\neg g$ are consequences of f. It is *decidable* [*entscheidungsdefinit*] if one of these two is always the case, i.e. if for every propositional function g, exactly one of g or $\neg g$ is a consequence of f. (2) *Not gabelbar.* An axiom system f is said to be *satisfied* if the proposition $(\exists R) f(R)$ holds (in the basic system). Given a function g, the system f is called *gabelbar* ("forkable") at g if both $f \& g$ and $f \& \neg g$ are satisfied. (Think for example of g as the axioms of parallels in Bolyai's "absolute" geometry, which is *gabelbar* at g because there are both Euclidean and non-Euclidean geometries.) An axiom system is said to be *gabelbar* if it is *gabelbar* at some g. If an axiom system is satisfied and not *gabelbar*, then any two models satisfy all the same propositional functions g. (3) *Monomorphic.* An axiom system is said to be *monomorphic* if it is satisfied and any two of its models are isomorphic. It is *polymorphic* if it has non-isomorphic models." [Awodey & Carus (1998), 15]

of conjunction), one obtains a propositional function. Hence, given an n-tuple of suitable variables of a proper type $\mathcal{R} = (r_1, r_2, \ldots r_n)$, one may express the whole axiom system as a single propositional function $f\mathcal{R} = f_1\mathcal{R}\&f_2\mathcal{R}\&\cdots\&f_m\mathcal{R}$, where $f_1, f_2, \ldots, f_m$ are the individual axioms. For example, consider the system of the Zermelo-Fraenkel axioms of set theory. Then the primitive symbols are $r_1 = \emptyset, r_2 = a, r_3 = b$, etc. (comprising only these three as individual constants, and in addition, a collection of variables), and the axiom system can be written as $\mathcal{R} = f_1\mathcal{R}\&f_2\mathcal{R}\&\cdots\&f_{10}\mathcal{R}$, where the f_i are the 10 axioms of the Zermelo-Fraenkel system ($f_1 =$ Axiom of Extensionality, $f_2 =$ Axiom of Foundation, $f_3 =$ Axiom of Subsets, etc.). A *model* of f is defined to be an n-tuple $A = (a_1, \ldots, a_n)$ of constants (of suitable types) from the basic system which satisfy f. Finally, a propositional function f is said to be *formal* if it respects isomorphism, i.e. if A and B are isomorphic structures of suitable type, and A satisfies f, then so does B. These remarks should suffice to make the following discussion comprehensible.

To work our way towards the *Gabelbarkeitssatz*, it is expedient to adduce the few theorems that figure in the proof. First, Carnap provides theorems about the pair of notions "monomorph" – "polymorph":

Theorem 1 *Daß ein Axiomensystem $f\mathcal{R}$ (mit q-stufiger Variabler)* MONOMORPH *ist, ist äquivalent mit:*

1. *zu f gehört genau eine Struktur, und zwar eine q-stufige, (nach Definition);*
2. *f ist erfüllt, und je zwei Modelle von f sind q-stufig isomorph:*

   $$(\exists)f\&(\mathcal{P}, \mathcal{Q})[(f\mathcal{P}\&f\mathcal{Q}) \rightarrow Ism_q(\mathcal{P}, \mathcal{Q})];$$

   Dies nehmen wir als DEFINITIONSFORMEL *für die Monomorphie*
3. *f ist erfüllt und nicht polymorph. [...]*

Theorem 2 *Daß ein Axiomensystem $f\mathcal{R}$ (mit q-stufiger Variabler)* POLYMORPH *ist, ist äquivalent mit:*

1. *zu f gehört mehr als eine Struktur (und zwar q-stufige Strukturen), (nach Definition);*
2. *es gibt zwei q-stufig nicht-isomorphe Modelle von f:*

   $$(\exists\mathcal{P}, \mathcal{Q})[(f\mathcal{P}\&f\mathcal{Q})\&\overline{Ism_q}(\mathcal{P}, \mathcal{Q})];$$

   Dies nehmen wir als DEFINITIONSFORMEL *für die Polymorphie*
3. *f ist erfüllt und nicht monomorph. [...]*

Here the expression $Ism_q(\mathcal{P}, \mathcal{Q})$ signifies an isomorphism (of degree q) between $\mathcal{P} and \mathcal{Q}$. Next we introduce the definition of the concept "*gabelbar*" within Carnap's system:

Definition 2 *$f\mathcal{R}$ sei "GABELBAR AN" $g\mathcal{R}$, wenn f mit g und mit $\overline{g}$ verträglich ist und g formal ist:*

$$(\exists)(f\&g)\&(\exists)(f\&\overline{g})\&(\mathcal{P},\mathcal{Q})[(g\mathcal{P}\&Ism_q(\mathcal{Q},\mathcal{P}))\rightarrow g\mathcal{Q}].$$

Das positive Kriterium dafür, daß f an g (k-)gabelbar ist, besteht somit in der Aufweisung eines gemeinsamen Modells für f und g, eines gemeinsamen Modells für f und $\overline{g}$ und eines Beweises für die Formalität von g.

Wir nennen $f\mathcal{R}$ "GABELBAR", wenn es ein $g\mathcal{R}$ gibt derart, daß f an g gabelbar ist. Das positive Kriterium für die (k-)gabelbarkeit von f besteht somit in der Aufweisung eines derartigen g.

Given these theorems and the definition of the notion "*gabelbar*", we may proceed to study Carnap's proof of the celebrated *Gabelbarkeitssatz* (I here reproduce Carnap's proof from [Carnap (2000)]. A sketch of the proof in English is given in [Awodey & Carus (1998), 16].) As it turns out, the proof does not establish the intended relationship between the fundamental notions involved. The attested proof hinges on inadequate and, strictly speaking, false assumptions that abrade the objective of Carnap, viz. establishing that the two notions of *semantic completeness* and the *categoricity* of a system of axioms are equivalent. The core of the difficulties that figure in Carnap's proof is that the notions of syntactic provability and truth in a system are not adequately demarcated from one another. This is not extraordinary, since the content of these concepts were not fully grasped by mathematicians or logicians at the time. Gödel's incompleteness results were announced at the Königsberg conference on epistemology of the exact sciences in September 1930. Carnap was among the first to assimilate their implications, and consequently abandoned his work in the field of general axiomatics. Moreover, Tarski's definition of truth for formalized languages was reported at a conference in Warsaw in 1931. The impact of this work on Carnap's thinking was equally notable. But neither of these works were available for Carnap when he first attempted to find an answer to the questions of general axiomatics that he and Fraenkel had formulated in 1927–1928. So, despite its central importance during the period 1927–1930, Carnap ultimately abandoned the *Axiomatics* project in early 1930. Incidentally, he did this almost exactly at the time that a summary of his Prague presentation appeared in *Erkenntnis*. [Carnap (1930d)]

It therefore came as no surprise that Carnap could not, given the improper conceptual tools that he had recourse to in devising the proof, attain the goal that he had set for himself. Let us now study the proof of the *Gabelbarkeitssatz* in detail, and see what the difficulties amounted to *in concreto*. After the exposition we are in a better position to evaluate the import of Carnap's ideas in this particular domain.

Carnap divides the proof of the *Gabelbarkeitssatz* in two parts (*Erster Teilsatz* and *Zweiter Teilsatz*), the first being the proof of the 'only if'-part:

Theorem 3 (*Der Gabelbarkeitssatz*) *Erster Teilsatz. Jedes polymorphe Axiomensystem ist gabelbar.*

Beweis. Das Axiomensystem $f\mathcal{R}$ sei polymorph. q bezeichne die Stufenzahl der Modellvariabeln $\mathcal{R}$; die Ausdrücke "Isomorphie", "Struktur", u. dergl. sind im folgenden immer im Sinne von "q-stufiger Isomorphie", "q-stufiger Struktur" usw. verstehen. Die Polymorphie bedeutet: es gibt zwei Modelle von f, etwa $\mathcal{R}_1$ und $\mathcal{R}_2$, die nicht isomorph sind; es gibt also:

$$\mathcal{R}_1 \tag{1}$$

$$\mathcal{R}_2 \tag{2}$$

$$\overline{Ism_q}(\mathcal{R}_1, \mathcal{R}_2) \tag{3}$$

Unsere Überlegung lautete nun: Da $\mathcal{R}_1$ und $\mathcal{R}_2$ nichtisomorph sind, so muß es eine strukturelle Eigenschaft geben, die dem $\mathcal{R}_1$ zukommt, dem $\mathcal{R}_2$ aber nicht. Die Aufgabe besteht jetz darin, eine solche Eigenschaft explicit anzugeben, und zwar in so allgemeiner Form, daß sie gemäß dieser Form in jedem konkreten Falle konkret konstituiert werden kann. Die Aufgabe ist einfach zu lösen; wir nehmen die Eigenschaft "die Struktur von $\mathcal{R}_1$ zu besitzen". Diese Eigenschaft kommt sicherlich dem $\mathcal{R}_1$ zu, dem $\mathcal{R}_2$ aber nicht, da ja $\mathcal{R}_2$ nicht isomorph mit $\mathcal{R}_1$ ist. Die genannte Eigenschaft können wir auch so ausdrücken: "isomorph mit $\mathcal{R}_1$ zun sein", als Aussagenfunktion: $Ism_q(\mathcal{R}_1, \mathcal{R}_2,)$; diese Aussagenfunktion sei mit $g\mathcal{R}$ bezeichnet:

$$g\mathcal{R} =_{Df} Ism_q(\mathcal{R}, \mathcal{R}_1) \tag{D}$$

Um f und g zu gabeln, müssen wir zunächst nachweisen, daß g formal ist. Gemäß (D) gilt tautologish:

$$(\mathcal{P}, \mathcal{Q})[g\mathcal{P} \to Ism_q(\mathcal{P}, \mathcal{R}_1)] \tag{4}$$

$(\sim)$

$$(\mathcal{P}, \mathcal{Q})[g\mathcal{P} \& Ism_q(\mathcal{Q}, \mathcal{P}) \to Ism_q(\mathcal{Q}, \mathcal{R}_1)] \tag{5}$$

$(\sim)$

$$(\mathcal{P}, \mathcal{Q})[g\mathcal{P} \& Ism_q(\mathcal{Q}, \mathcal{P}) \to q\mathcal{Q}] \tag{6}$$

Dies besagt, daß g formal ist. Nach [(*)]:[104]

$$Ism_q(\mathcal{R}_1, \mathcal{R}_1)$$

$(\sim)$ (nach D)

$$g\mathcal{R}_1 \tag{7}$$

Aus (3) (nach [(**)][105]):

$$\overline{Ism_q}(\mathcal{R}_2, \mathcal{R}_1);$$

[104](*) Die q-stufige Isomorphie ist stets (d. h. für einen beliebigen Wert von q) (total) reflexiv:

$$(P, q) Ism_q(P, P).$$

[105](**) Die q-stufige Isomorphie ist stets symmetrisch:

$$(P, Q, q)[Ism_q(P, Q) \to Ism_q(Q, P)]$$

$(\sim)$ (nach D)

$$\overline{g}\mathcal{R}_2 \tag{8}$$

Aus (1) und (7):

$$f\mathcal{R}_1 \& g\mathcal{R}_1 \tag{9}$$

Dies besagt, daß f mit g verträglich ist. Aus (2) und (8):

$$f\mathcal{R}_2 \& \overline{g}\mathcal{R}_2 \tag{10}$$

Dies besagt, daß f und $\overline{g}$ verträglich ist. Also ist f an g gabelbar.

This proof constitutes the demonstration of the necessity of the property of "gabelbarkeit" for the non-categoricity of an axiom system. (Note that the statement is equivalent with the updated formulation "The categoricity of an axiomatic theory implies its semantic completeness.") Carnap then proceeds to prove its converse, i.e. the theorem

Theorem 4 (*Der Gabelbarkeitssatz*) *Zweiter Teilsatz. Jedes gabelbare Axiomensysteme ist polymorph.*

Beweis. Das Axiomensystem $f\mathcal{R}$ sei gabelbar. Das bedeutet: es gibt eine Aussagenfunktion $g\mathcal{R}$ mit derselben Variabeln derart, daß (1) g formal ist, (2) f mit g verträglich ist, (3) f mit $\overline{g}$ verträglich ist; in Formeln:

$$(\mathcal{P},\mathcal{Q})[g\mathcal{P} \& Ism_q(\mathcal{P},\mathcal{Q}) \rightarrow g\mathcal{Q}] \tag{1}$$

$$(\exists \mathcal{G})(f\mathcal{G} \& g\mathcal{G}) \tag{2}$$

$$(\exists \mathcal{Z})(f\mathcal{Z} \& \overline{g}\mathcal{Z}) \tag{3}$$

Aus (2) und (3):

$$(\exists \mathcal{G},\mathcal{Z})(f\mathcal{G} \& f\mathcal{Z} \& g\mathcal{G} \& \overline{g}\mathcal{Z}) \tag{4}$$

Aus (1) [...]:

$$(\mathcal{P},\mathcal{Q})[g\mathcal{P} \& \overline{g}\mathcal{Q} \rightarrow \overline{Ism_q}(\mathcal{P},\mathcal{Q})] \tag{5}$$

Aus (4) und (5):

$$(\exists \mathcal{G},\mathcal{Z})[f\mathcal{G} \& f\mathcal{Z} \& \overline{Ism_q}(\mathcal{G},\mathcal{Z})] \tag{6}$$

Dies besagt: f ist polymorph.

Carnap has thus proven the following theorem:

Theorem 5 [***Der Gabelbarkeitssatz***]

*Die Begriffe "*POLYMORPH*" und "*GABELBAR*" sind äquivalent.* □

There are several points to be made regarding the proof. To begin with, the fundamental shortcoming in the method that Carnap used lay in its insensitivity to the distinction between the axiom system under investigation and the logical language in which those investigations were conducted. The understanding that has been gained with the help of historical hindsight is that the reason why Carnap found these ideas natural is that he held at the time a "universalist" conception of language. (This theme will come up in more depth later.) According to this conception, there is only one logic, and all logical analyses of and statements about an axiom system are to be conducted in the universal system of which the axiom system under investigation is a part. Thus, Carnap did not at this stage have the logical means to distinguish between the question of provability in an axiomatic system and the question of provability of a statement *about* axiom systems. This is tantamount to not having the distinction between metalanguage and object language. Another problem that stems from the universalist conception of language is that according to Carnap's idea of the basic system, the *Grunddisziplin*, the sentences of that system have 'content' (i.e. they are interpreted [*inhaltlich*]). This implies that the notion of provability in a purely syntactic sense is impossible; indeed, we have not provability but an "absolute" concept of truth. Therefore it becomes possible for Carnap to pass from statements of the form

$$\text{for no } g, f \to g \text{ and } f \to \neg g$$

to

$$\neg(\exists g)(f \to g \& f \to \neg g) \text{ holds.}$$

This means that "not-$\vdash p$" and "$\vdash \neg p$" are synonymous, and thus interchangeable. It is the combination of these two assumptions (universality of the single system and absolute truth within that system) that jeopardizes Carnap's goal of elucidating the notion of provability and completeness. The issue is not so much that Carnap's results are false (indeed, with the interpretation of the notions given by Carnap they turn out to be for the most part trivially true), rather than the fact that given the tools he applied, he was not addressing the problems he wished to consider.

It is expedient to consider the issue from a more modern point of view and state the problems Carnap *intended* to tackle using updated terminology. The three notions of completeness that formed the object of Carnap's and Fraenkel's investigations in the late 1920s can be related by the following conjectures (formulated as theorems):

Theorem I: An axiomatic system S is consistent (no contradiction is deducible from it) if and only if it is satisfiable, i.e., has a model.

Theorem II: An axiomatic system S is semantically complete (non-forkable) if and only if it is categorical (monomorphic).

Theorem III: An axiomatic system S is deductively complete if and only if it is semantically complete (non-forkable). [Reck (2007) [Friedman & Creath (2007), 187]]

To correlate these with Carnap's statements in the *Untersuchungen*, we may indicate that Theorem I corresponds to Carnap's "Satz 2.4.9" [Carnap (2000), 100], Theorem II (which is, of course, the *Gabelbarkeitssatz*) to "Satz 3.4.10" [*ibid.*, 138] and Theorem III to "Satz 3.6.1" [*ibid.*, 144]. Remember that Carnap is working in higher-order logic, not first-order logic. It is then immediately clear that Theorem I and Theorem III fail to hold.[106] But Theorem II is an interesting case that merits further scrutiny. This will become apparent in a moment. For now I would like to delve in a little more detail to the sources of the difficulties and ambiguities that figure in Carnap's exposition.

Carnap's approach in the *Untersuchungen* might be properly described as Russellian rather than metamathematical. The distinction between the system which is under investigation (object language) and the system in which these investigations are effected (metalanguage) is not yet made. This means that notions such as "logical consequence", "consistency", and so on, are explicated *internally* to the logical system under consideration. To provide an example, consider the notion of logical consequence. Carnap's definition of "logical consequence" (deducibility in Carnap's general sense and a direct descendant of Russell's notion of "formal implication") contains an ambiguity that undermines the entire project of the *Untersuchungen*:

Definition 3 *The proposition* $Q(t_1, \ldots, t_n)$ *is a* logical consequence *of the proposition* $P(t_1, \ldots, t_n)$ *if and only if* $\forall x_1 \ldots \forall x_n (P(x_1, \ldots, x_n) \supset Q(x_1, \ldots, x_n))$ *holds.*[107]

The ambiguity here derives from the ambiguity of the verb "holds". There are two different senses of "holding" that might intelligibly figure here. The first one attaches to the verb the interpretation "being deducible in the given formal system." This is, effectively, the notion of syntactic consequence in higher-order logic. The second one attaches to the verb the interpretation along the lines of "being true" (remember that Carnap is working in a universal domain, following Frege and Russell). Then the meaning of the definition is close to the contemporary one of the notion of higher-order semantic consequence. It is now pertinent to ask, which of these two interpretations is Carnap working with in the *Untersuchungen*? What does he mean, in particular, when he talks about "deducibility"? Erich Reck has summarized the difficulties that the reading of *Untersuchungen* has posed for modern scholars:

[106] Reck [*ibid.*] gives the following lucid demonstration of this: "Let *PA* be the higher-order Dedekind-Peano axioms (assumed to be consistent); let G be the sentence shown to be true but not provable from *PA* in Gödel's Incompleteness Theorem. Then *PA* together with $\sim G$ is consistent but not satisfiable. This shows that the 'only if'-part of Theorem I fails. As neither G nor $\sim G$ is provable, *PA* is not deductively complete,; but it is semantically complete, because categorical. That shows that the ['if'-part (* Reck erroneously writes " 'only if' part")] of Theorem III fails. [...]"

[107] Note that the operation of quantifying out the constants $t_1, \ldots, t_n$ is effectively similar to the now standard idea of varying the interpretation of all the non-logical symbols in the language.

> From a contemporary point of view, one would expect him to work with syntactic consequence, especially since that seems to be the notion built into deductive completeness as used in Theorem 3. Recall also Fraenkel's informal characterization of deductive completeness in *Einleitung* (1928) [...] which Carnap seems to want to explicate. Similarly, one would expect syntactic consequence to be built into Carnap's notion of consistency as occurring in Theorem 1. Overall, however, Carnap leans more towards semantic notions in *Allgemeine Axiomatik*, which point in the direction of semantic consequence; and in so far as this is the case, his explications of Fraenkel's distinctions are not adequate, especially that of deductive completeness. But most importantly, Carnap simply does not seem to be clear about the difference between syntactic and semantic consequence, both of which he can be read as invoking, at different points in his discussion, as if they were equivalent. [...] In other words, he is implicitly working with an inchoate amalgam of the two notions, and this is directly affecting his understanding of Theorems 1 and 3. [Reck (2007), 189]

Similarly, other important definitions he provides are also simply inadequate for his purposes. Another example is the notion of "consistency" which Carnap defines for a finite system of axioms as follows (notation updated [*ibid., 190*]):

Definition 4 *Suppose that $P(t_1, \dots, t_n)$ is the conjunction of the given axioms. Then the axiom system is called* consistent *if and only if $\neg\exists Q \forall x_1 \dots \forall x_n (P(x_1, \dots, x_n) \supset (Q(x_1, \dots, x_n) \wedge \neg Q(x_1, \dots, x_n)))$ "holds".[108] [Carnap (2000) [1927–1930], 97]*

Besides thes's the specific shortcomings in the *Untersuchungen*, the *Hauptschwierigkeit* of the attempt at a "General Axiomatics" lay in the explicit (albeit unreflected) attachment to a Fregean and Russellian "universalist" conception of logic. This is manifest in the way how Carnap uses a single formal system embedded in a fixed, all-encompassing background language. This framework is endowed with a system of fixed rules of inference that govern the logical reasoning taking place in that system. This universal framework does not allow the variation of the interpretations of the symbols of the language, indeed, the whole notion of interpretation is something intangible in this context. The symbols denote the objects or relations they do by the sole means of their specific roles within the universal system. The distinction between an object language and a metalinguistic framework is not made. But then, this distinction is exactly what from our modern point of view is seen to be essential for defining the notions and the relations between these notions that Carnap is interested in, namely, the several notions of completeness, consistency and so forth.[109] That Carnap was deprived of such a distinction that allows for the

[108] N.B. the same ambiguity as in the definition of "logical consequence".

[109] This view can be contested, of course. The metalinguistic perspective is not essential for defining all the notions traditionally conceived as *metalogical*. Jaakko Hintikka has shown how the notion of truth, for example, can be defined *within* a language endowed with suitable properties (the most prominent examples being the so-called IF-languages in Hintikka's locution). [Hintikka (1989), (1991)]

definitions of the modern notions of syntactic and semantic consequence, can be seen, in large part, as a reason for the failure of his project.

An index of the novelty of the whole domain of questions that Carnap was tackling with was the way in which the faults of his exposition escaped the notice of the most prominent mathematicians and logicians of the period. Indeed, Carnap had shown his manuscript for part I to Fraenkel and Gödel already in 1928. Moreover, he was presenting the material in his public lectures (as mentioned above) and in talks in Vienna that same year, and then, most importantly, at a conference in Prague, "the First Conference on Epistemology of the Exact Sciences", in 1929. In this conference, in particular, the immediate response to Carnap's work was enthusiastic, as is witnessed by a corresponding diary entry written shortly afterwards:

> My lecture: Investigations in General Axiomatics; just a brief summary. But the proof is requested, and acknowledged. Though it was late, a lively discussion on the basic issues afterwards; von Neumann, Zermelo, Hahn, Fraenkel said that a final judgement will only be possible when the complete proof [especially of Theorem 2 and 3 above] is available. Amazing interest in my Investigations. [Quoted in Awoey & Carus (2001), 162]

Furthermore, Gödel's understanding of Carnap's project is reflected in the third paragraph of his dissertation 1929, where he says that: "if we replace the notion of logical consequence [...] by implication in Russell's sense", then the assertion of completeness of the logical system is provable in a few straightforward steps. He credits Carnap for this observation. [Gödel, *Collected Works*, vol. I, 63] But what he does not note is that at the time (explicitly in the *Untersuchungen*), Carnap thinks that logical consequence *is* formal implication in the sense pointed to above. The tides began to turn in 1930 when Tarski visited Vienna, and the two had several conversations together. As a result of these discussions, which touched also the topic at hand, Carnap began to doubt the adequacy of his approach. This is reflected in his diary remark from that period:

> Tarski visits me [...] talked about my *Axiomatik*. It seems correct, but certain concepts don't capture what is intended; they must be defined metamathematically rather than mathematically. [Quoted in Awodey & Carus (2001), 163]

Another impulse to re-evaluate the views exposed in the *Untersuchungen* came from discussion with Kurt Gödel, who was, at the time, a research student at the University of Vienna. Gödel attended Carnap's talks and classes (as evinced by Hao Wang [Wang 1987]) and the two had many discussion upon logic in general and Carnap's project of "General Axiomatics" in particular. As is now well known, at least partly influenced by his prolonged contact with Carnap and partly by the work of Hilbert and his school at Göttingen (foremost by the important *Grundzüge der theoretischen Logik* (1928) by Hilbert and Ackermann), Gödel discovered his now celebrated results: the Completeness Theorem for first-order logic and the Incompleteness Theorems for arithmetic and higher-order logic.[110] As Warren Goldfarb reports, "Carnap was among the first of those whom Gödel told of his incompleteness discovery, in a conversation on 26 August 1930, and he was in the audience at Gödel's first public announcements of his result during a conference in Königsberg on 5–7 September 1930" [Goldfarb (2003) [Gödel, *Collected Works*, vol. IV, 335]] Carnap was, then, quick to notice the shortcomings of his proposals for proofs pertaining to the relations of the central metamathematical notions. Consequently, he abandoned the research program of "General Axiomatics" and turned to investigating the possibilities of devising a metalogic on another basis. However, this contingent turn of events left an important aspect of Carnap's work undiscovered. Only recently, due to the work of Steve Awodey and Erich Reck [Awodey & Reck (2002a)], this aspect has been brought back to light.

Thus, considering Carnap's contributions to metamathematics afresh, what can be said about the *Gabelbarkeitssatz* from our modern vantage point? We happen to know (due to the result of Lindebaum and Tarski (1935)) that the 'if'-part is actually true (i.e. the categoricity of an axiomatic theory implies its semantic completeness). This result holds, not only in the first-order predicate logic, but also in higher-order logic for axiomatic theories. The question about whether the converse is true is to this date undecided, at least in its most general formulation. As Awodey and Reck remark, the "inference from semantic completeness to categoricity, depends crucially on two background conditions: first, it depends on the logical language used, in particular on what sorts of sentences φ are supposed to occur in the definition of semantic completeness. Clearly, the inference fails, e.g., if we restrict attention to just first-order sentences. But what about the case of higher-order logic? Here, secondly, it is crucial to be precise about what is meant by 'axiomatic theory'. Indeed, it is not hard to see that the inference from semantic completeness to categoricity fails again if we consider general 'theories' in the sense of arbitrary sets of sentences in some given language." [Awodey & Reck (2002a), 25] Given these reservations it then becomes possible to address the following question that sharpens the original problem considerably: "[F]or a theory $\mathbb{T}$ with finitely many axioms in higher-order logic, does the semantic completeness of $\mathbb{T}$ imply its categoricity?" [*ibid.*] Answering this question was one of the main tasks that Carnap set himself in the second half of the 1920s (but in which he failed, as we have seen). We have come to recognize now the intrinsic interest of this and related questions that Carnap tackled in his *Untersuchungen* and research papers from that period. As Awodey and Reck have convincingly shown, the insights gained by Carnap in the 1920s, when freed from the restrictions that accompany them, open up fruitful vistas for further development in the foundations of mathematics and in the study of higher-order axiomatics.[111] This makes Carnap's work in logic appear in a completely new light and testifies to his importance as a pioneer of mathematical logic. There is, however, another important point to be noticed. Notwithstanding their interest from the point of view of foundational questions in mathematics, the ideas in the

[110] In 1975, in an answer to an inquiry by Burke Grandjean, then a doctoral student of sociology at the University of Texas, about the important influences on his work on completeness, Gödel picked out *Hilbert and Ackermann (1928)* and Carnap's lectures. [Gödel, *Collected Works*, vol. IV, 449]

[111] Cf. especially the sequel to [Awodey & Reck (2002a)], viz. [Awodey & Reck (2002b)].

Untersuchungen crucially prepared Carnap for the next step in his escape from the highly restrictive framework of rational reconstruction that was essentially facing a crisis. Carnap took the next step quite soon, and the result was one of the most important books of twentieth-century philosophy, *Logische Syntax der Sprache*.

Chapter 5

Logical Syntax and Semantics as Vehicles for Analysis

5.1 The Phase of Liberation

5.1.1 *Der Wiener Denkstil*

The title of this subsection refers to a little book by Arnold Keyserling, a compilation of seminal writings by Ernst Mach, Rudolf Carnap and Ludwig Wittgenstein, the purpose of which is to illustrate and render better understood the singularly original mode of philosophical thinking that is foremost associated with the Vienna Circle, but which derives its basic *ethos* from the general intellectual atmosphere of *fin-de-siècle* Vienna. The quite conspicuous differences in their overall approaches and vocabulary notwithstanding, each of these three philosophers attempts to promote a new approach to philosophical problems in general – or, alternatively, novel means of *avoiding* certain philosophical problems in particular. The approach can most naturally be described as *antimetaphysical*. The selection from Mach in Keyserling's book is "Antimetaphysischen Vorbemerkungen", from Carnap, "Die Aufgabe der Wissenschaftslogik", and from Wittgenstein, extracts from "Philosophische Untersuchungen". Keyserling attempts to link the ideas of all these thinkers under a broad intellectual approach that he calls *Der Wiener Denkstil*. What does he mean by this locution? In the introductory part of his book he discerns three important concepts around which this particular intellectual mode of thought settles. These concepts are those of *element*, *structure* and *game*. These are the fixed points of the intellectual ambit of a whole series of philosophers representing a particularly Viennese style of thinking. As Keyserling puts it himself:

> Dieses Wiener Denken kreist um drei komplementäre Begriffe: um die Bedeutung des *Elements*, der *Stuktur* und des *Spiels*. Dies ist nun eine ganz

> neue und eigenartige Trinität, die in dieser Form in der europäischen Geistesgeschichte geworden war; sonst bewegte sich das Denken im Rahmen allgemein geläufiger Gegensatzpaare, wie Ding und Erscheinung, Idealismus und Materialismus, Notwendigkeit und Freiheit, Individuum und Gemeinschaft. Doch in Wien wurde zum erstenmal eine ganz andere Problematik angegangen, die in folgender These gipfelt: alles philosophische Denken vollzieht sich zwischen den drei Polen des klar bestimmten Elements, einer festgefügten Struktur von Gesetezen und deren möglicher freier Kombination; so wie sich zum Beispiel alle materiellen Körper als Teil oder Verbindung der gleichen chemischen Elemente fassen lassen, und ihre Verknüpfung ganz bestimmten Gesetzen folgt; oder wie sich die Sprache aus den immer gleichen Buchstaben und Satz-Bausteinen (Elementen), nach gewiseen Gesetzen (Struktur) zusammengefügt, zu freier Aussage (Spiel) aufbaut; oder wiederum die Musik auf der Abwandlung einer begrenzten Anzahl Tönen zufolge den Gesetzen der Harmonik beruht; und wie sich schließlich selbst die individuelle Charakteristik einer menschlichen Persönlichkeit auf die besondere Kombination der Gene in den Chromosomen, also auf die besondere Verknüpfung der Bausteine der Erbanlage, zurückführen läßt.
>
> Diese Gedanken bewegten nun un Wien eine ganze Reihe von Geistern, in deren Werk sich die drei Begriffswelten Element, Struktur, Spiel immer klarer und wesentlicher herausschälten. Wir wollen versuchen, der Gestaltwerdung dieser Problematik bei drei Philosophen nachzugehen: bei *Ernst Mach*, *Rudolf Carnap* und *Ludwig Wittgenstein*, und anschließend die Folgen ihrer Entdeckung aufzeigen. [Keyserling (1965), 5–6]

I think this way of approaching the thoughts of these three thinkers is illuminating. It provides, particularly with respect to Carnap and Wittgenstein, a conceptual framework in which to assess their work in a broad perspective, enabling one to discern the details of their intellectual relationship and the transitions in their thought. Especially the problematic relation between Carnap and Wittgenstein becomes more comprehensible. Although they shared a bundle of important basic notions and presuppositions, their relationship (in both collegial and personal terms) was fraught with bitter controversy (more often than not instigated by Wittgenstein's flamboyant character). The tension between the two becomes most pronounced in connection with the issue of physicalism. I will comment on this in the section **5.1.5**. There it will also become clear how their shared conviction of the superfluity of metaphysics figured in that debate. Now it is expedient to try to understand how Carnap attempted to answer the challenge set by the crisis of the program of rational reconstruction alluded to in the section **4.9.7**. I will now turn to the crucial insight that launched Carnap onto the path that resulted in a formulation of a metalogic and its polished version in the book *Logische Syntax der Sprache*.

5.1.2 The path to metalogic

The story of the process by which Carnap came to formulate a new foundation for his program of reconstruction is made more comprehensible if it is embedded in the framework of universalism. This term alludes to a general conception of language that already Frege entertained in his *Begriffschrift* in 1879. Frege's system in that book was intended to represent a universal language the purpose of which was to elucidate the laws of rational thought. This attitude towards language as a medium of thought and rational discourse has been coined as "the universal conception of logic and language" by Jean van Heijenoort. The essential characteristics of this conception comprise the theses that nothing can be said or expressed outside the system itself and that there does not exists a meta-level on which one could discuss the meta-systematic or metalogical questions concerning the original (not necessarily formal) system. These meta-systematic questions include, *inter alia*, the question about the consistency of the system, the one concerning the independence of the axioms of the system, and the one concerning the completeness of the system. The most eminent philosophers in the analytic tradition have followed Frege in propounding this universalist conception of language. But which features of Frege's thought on language actually exemplified such universalism? According to van Heijenoort:

> [The] universality of Frege's *lingua characterica* is, first, the universality that quantification theory has in its vocabulary and that the propositional calculus lacks. Frege frequently calls Boole's logic an 'abstract logic', and what he means by that is that in this logic the proposition remains unanalyzed. the proposition is reduced to a mere truth value. With the introduction of predicate letters, variables, and quantifiers, the proposition becomes articulated and can express a meaning. The new notation allows the symbolic rewriting of whole tracts of scientific knowledge, perhaps all of it, a task that is altogether beyond the reach of propositional calculus. We now have a *lingua*, not simply a calculus.
>
> [...] However, the opposition between *calculus raticinator* and *lingua characterica* goes much beyond the distinction between the propositional calculus and quantification theory. The universality of logic expresses itself in an important feature of Frege's system. In that system the quantifiers binding individual variables range over all objects. [...] Frege's universe consists of all that there is, and it is fixed.
>
> [...] Another important consequence of the universality of logic is that nothing can be, or has to be, said outside of the system. And, in fact, Frege never raises any metasystematic questions (consistency, independence of axioms, completeness). Frege is indeed fully aware that any formal system requires rules that are not expressed in the system; but these rules are void of any intuitive logic; they are 'rules for the use of our signs.' In such a manipulation of signs, from which any argumentative logic has been squeezed out, Frege sees precisely the advantage of a formal system. [van Heijenoort (1967) [Hintikka

(1997), 233–239]]

More generally, we may present the differences between these two conceptions schematically in a table as follows:[1]

UNIVERSALITY OF LOGIC/LANGUAGE	THE MODEL-THEORETIC TRADITION
1. Interpretation cannot be varied.	1. Interpretation can be varied.
2. Model theory is impossible (or irrelevant).	2. Model theory is possible (and important).
3. Only one world can be talked about.	3. Many worlds can be talked about.
4. One domain of quantification in the last analysis.	4. Ranges of fully analysed quantifiers can be different.
5. Logical truths are about this (the actual) world.	5. Logical truth as truth in all possible worlds.
6. (Genuine) metalanguage is impossible.	6. Metalanguage is possible and legitimate.
7. Truth as correspondence is unexplainable/unintelligible.	7. Truth as correspondence is a legitimate notion.

Now, the development of Carnap's philosophical views — especially those concerning language — can certainly be investigated within this framework. Although it is too coarse-grained to function as an overall interpretive tool in studying Carnap's philosophical development, I will briefly indicate how it can be understood to shed some light on Carnap's philosophy. Given van Heijenoort's distinction between *Lingua Universalis* vs. *Calculus Ratiocinator*, it is then possible to demarcate, roughly, two phases in the development of Carnap's thought. The first phase comprises the work that Carnap did while in Vienna at the time when the Schlick Circle gathered regularly, i.e., during the years 1926–1931. During that time Wittgenstein was a major influence for the members of the Circle, attracting most prominently the attention of Moritz Schlick himself and Friedrich Waismann. While more reserved in his appraisal for the new and radical ideas of Wittgenstein, Carnap adopted a notable proportion of them into his own thinking, as we have seen. It is precisely the *Tractarian* idea of a uniform, all-encompassing linguistic framework that underlies Carnap's idea of rational reconstruction, so central to the Vienna Circle, of which the *Aufbau* had been a paradigmatic example. This phase can most naturally be thought of as one during which Carnap adhered to a purely universalist conception of language. It is the gradual loosening of this strict adherence to a single uniform logical framework, *Lingua Universalis*, that finally amounts to the abandonment of the restrictive conception of a single language of science, resulting in a transition to another phase. During this new phase Carnap comes to hold a representation-theoretical view of language in line with van Heijenoort's idea of a *Calculus Ratiocinator*. There are now many possible linguistic frameworks from which to choose according to the analytic task at hand. The choice among them is no longer a question of picking a *correct*

[1]Adapted from [Kusch (1989), 6–7]

language for scientific purposes but rather a question about pragmatic utility and fruitfulness. In this sense there definitely is a transition in Carnap's philosophy the particular characteristics of which the distinction of van Heijenoort undoubtedly captures. Accommodating this qualitative change within its purview it sheds some light on the peculiarities of Carnap's philosophical development. However, the framework is too rough for assessing the multilayered character of Carnap's thinking. Indeed, elements of the pragmatic ingredients in Carnap's thinking are already visible in his early work on geometry. In particular, in *Der Raum* Carnap discussed explicitly the necessity of adopting different linguistic frameworks to make sense of the different modes in which philosophers and scientists talk about space and spatial perception. The seeds of linguistic plurality are there right from the beginning, and the overtly universalist conception of language that dominated the outlook of Vienna Circle in the late 1920s has to be seen more as a sign of a solidification of particular aspects of Carnap's thought than as a fundamental, all-encompassing transformation. This transition is an index of the continual efforts of Carnap to increase the precision of the linguistic tools and forms of expression that he has recourse to in expressing the content of his philosophical views. Under this view, the universalist element in Carnap's thinking reflects only the range of linguistic forms that Carnap at the time conceived as capable of expressing that content with sufficient rigour. As soon as one recognizes that a constant feature of Carnap's philosophical development is the gradual extension of the scope of the problems that can be tackled with precision and exactness, the universalist conception that he entertains at one stage becomes an understandable link in the overall development of his thought. One to emphasize this aspect in Carnap's thinking was Abner Shimony. In his reminiscences of Carnap he put this point succinctly:

> I do not want to suggest, however, in pointing to the systematic character of Carnap's thought, that he was inflexible. [...] I want to add that he took particular delight in technical advances which permitted him to widen the scope of his investigations without loss of precision. Perhaps the most important such advance for him was Tarski's work on the concept of truth, which enabled him to extend his investigations of language from syntax to semantics. In *Testability and Meaning* and *Foundations of Logic and Mathematics* he himself did the main technical work which permitted the relaxation and extension of the empiricist criterion of meaningfulness. A third and more recent example was Kemeny's work on models, which permitted him to define *c*-functions on languages which could not be readily analyzed semantically in terms of state-descriptions. The general drift of his philosophical development towards greater flexibility, openness, and richness indicates that his early program for the elimination of metaphysics was never an attempt to contract the scale of the world or to view it anthropocentrically, as some critics have claimed. Rather, it was the result of his intense desire to understand things clearly, and my personal impression was that he experienced much joy in finding that clarity is compatible with greater and greater scope. [Shimony (1975) [Hintikka (1975b), XXVIII–XXIX]]

The gist of Carnap's work in philosophy in general, and logic in particular, consists of the attempt to delineate the expressive power of language (more accurately, formal languages) and to make explicit the *structural features* of various disciplines that are revealed through the expressive function of (formal) language.[2] In this way his philosophical agenda is seen to comprise a bewildering group of problems. In the most general terms, Carnap's program to define the boundaries of meaningful discourse about logic, language and truth was essentially built on the foundations of the classical questions pertaining to philosophy of logic that already Plato had addressed. These questions were:

1. What is it that can properly be called true or false?
2. What link is it that makes valid inference possible, or what is necessary connexion?
3. What is the nature of definition and what is it that we define? [Kneale & Kneale (1962), 17]

Certainly, the answers that Plato gave to these questions do not touch exactly the same issues that Carnap and other modern philosophers or logicians were interested in. This has undoubtedly much to do with the fact that for Plato these questions were to a marked degree obscured by the metaphysical and epistemological questions that formed the true focus of his philosophical enquiries. Nevertheless, the questions themselves originate with him, and jointly with them, philosophy of logic. In any case, they constitute a general framework in which Carnap's work in philosophy may be assessed. Indeed, during the years 1931–1947, Carnap's work very much focuses on these questions, or their closest relatives.

The vision of a "general theory of linguistic forms"

In the summer of 1931, Carnap moved to Prague to take up a professorhip at the German University. During this period his philosophical thinking went through a transformation the repercussions of which would affect everything that he wrote henceforth. During his first years at Prague, Carnap was principally occupied with writing his *Logische Syntax der Sprache* [Carnap (1934)], the idea of which had come to him, when, as he later recounted,

> the whole theory of language structure and its possible applications in philosophy came to me like a vision during a sleepless night in January 1931, when I

[2] Herbert G. Bohnert, in reporting his experience of attending a seminar conducted by Carnap, gives a succinct description of the amazement he felt after having conceived the startling consequences of Carnap's overall view: "Carnap went on to say, after supplying the needed details of definition, that the importance and power of logicism stem from the fact that all inferences about any subject matter depended only on the relational structure, and structural properties, of that subject matter. Completely corresponding results were assured for quite different subject-matters provided only that there be a structural similarity. And — my thoughts echoed — all structures and structural properties are definable by logic alone. We seemed in possession, then, of all possible forms of all possible knowledge!" [Bohnert (1975)[Hintikka (1975b), XXXVII–XXXVIII]]

> was ill. On the following day, still in bed with a fever, I wrote down my ideas on forty-four pages under the title 'Attempt at a metalogic'. [Carnap (1963), 53]

He went on to characterize his technical end as follows:

> One of my aims was to make the metalanguage more precise, so that an exact conceptual system for metalogic could be constructed in it. Whereas Hilbert intended his metamathematics only for the special purpose of proving the consistency of a mathematical system formulated in the object language, I aimed at the construction of a general theory of linguistic forms. [*ibid.*]

Inspired by his *Versuch* Carnap intended within the following two weeks to construct a "suitable meta-language", or a canonical "meta-logic". This labour resulted in a formulation of a system of language that was soon to become known as the "formal mode of speech". However, despite the intensive and inspired elaboration on the *Versuch*, the first version of the system turned out to be insufficient and it took Carnap a considerable effort to work out modifications of this system that would ultimately constitute an acceptable formulation, resulting in the Language I of the *Logische Syntax der Sprache* published in 1934. The results of the initial formulation of the new approach were to be presented in an elementary exposition of logic and method (as a part of a popular series of monographs that Neurath was planning) tentatively entitled *Introduction to Scientific Philosophy* ('Einführung in die wissenschaftliche Philosophie' preparing the publication of *Die Aufgabe der Wissenschaftslogik* (1934) within Neurath's series *Einheitswissenschaft*). As can be read in the UCLA papers [UCLA (1930c), (1931a), (1931c)] its first part was intended to be an overview of language, logic and mathematics entitled "The Language of Science". The second part was intended to include a treatment of empirical sciences, including psychology and social sciences, naturally, with a view to their underlying unity. Carnap began this project in late February and, by the middle of March (actually 11 March), had written 115 pages in the Stolze-Schrey shorthand (habitually used by him).

What was the general goal of this project? Quite simply, to make the following view both coherent and plausible: that all of logic — logic in the sense of inferential practices and reasoning — could be represented purely formally. This implied that logic was to be constructed from the bottom up without recourse to semantics, i.e. with no machinery for exemplifying the referential relations between the terms of the formal system and some extralogical domain. Carnap envisioned here a system that was in a sense a realization of the Leibnizian idea of *calculus ratiocinator*, a purely formal system of rules without interpretation. The starting point for this program is Carnap's method of introducing elementary propositional logic using truth tables. Thus Carnap proceeds to define the connectives, the notions of "range" and "content", logical consequence, and the concepts of tautology and contradiction. Of course, the notion of tautology that was central to the conception of logic originated by Wittgenstein could not be defined without truth tables (in the context of propositional logic), because it is an inherently *semantical* notion.[3]). But then Carnap abruptly changes the mode of speech, laying emphasis on the novelty of the approach:

> Now we shall go on to revisit this entire train of thought in a different form, this time with the help of the '*calculus*', of calculation with logical formulas. We must learn this method as it is an exact and indispensable tool for applications ... In a sense, we have two languages before us here. The first consists of the 'logical formulas', sign-complexes in which the signs 'p', 'q', ..., '$\sim$', '$\vee$' are concatenated in a certain way. In the second language, which we are now considering, we pay no attention to the meaning [*Sinn*] of the logical formulas; we do not distinguish, as in the first language, whether p is true or false. In this second language, the so-called *metalogical language*, we form '*metalogical' concepts* with which to describe the formulas of the first language as figural complexes.[4] [UCLA (1931e), 17–18]

The particular challenge that arises here is the definition of the inherently semantical notions of "tautology" and "content" in purely formal terms. It is interesting to note that while Carnap is not able entirely to accomplish this in the first part of "The Language of Science", he comes in the process to use concepts that were to become essential for him later in his 'semantical phase'. Thus, the notion of the content of a sentence is defined along Wittgensteinian lines as the set of rows of its truth table a sentence excludes, the 'range' of the sentence being defined as the complement of "content", i.e. the set of rows in which the sentence is true. The problem of defining this pair of concepts in a purely formal manner merged in a more general problem of defining the interrelations of them, since tautology as a sentence that is true in *all* rows of its truth table had *no* content at all. These stipulations are evident precursors of the definitions of the concept of semantic information worked out by Popper, Carnap and Bar-Hillel in the 1950s. What the stipulations intended to capture was a sense in which metalogic would be provided with a criterion of 'pure formality', i.e. criteria to stipulate [*festzusetzen*] whether a given formula is a tautology. Now, these stipulations do not entirely dispense with the notion of "truth", as can be seen. Although the implicitly semantical notions of "tautology" and "content" are here defined by the purely syntactic notion of "set of rows in the truth table of a given sentence", the assignment of the truth values to the constituents of different sentences is not an entirely arbitrary matter. Even if these assignments are thought of as linguistic conventions, they do not repudiate the inherently semantical nature of the truth-value ascriptions. It might well have seemed possible in Carnap's eyes to circumvent this semantic ingredient by having recourse to this syntactical way of using the truth tables. However, the truth tables were soon found out to be insufficient for the task of

[3]According to the classical division of levels in the study of language, there are three main categories of inquiry (summarized as follows in [Carus (2007), 21]): (1) *syntax* considers languages as pure calculi, *in isolation from anything extralinguistic they might be thought of representing*; (2) *semantics* considers languages as representing extra-linguistic affairs, but still in isolation from actual uses of language by humans; and (3) *pragmatics* considers languages in relation to their use contexts and their users.

[4]Cited in [Carus (2007), 234].

defining the tautologies which in turn formed the backbone of the idea of purely formal relations between sentences. Gödel's recent address of the problem of the *unentscheidbare Sätze* in *Principia Mathematica* and the accompanying proof of the incompleteness theorem resulted in a dilemma: if the property of tautologousness were taken as a the principal criterion of 'formality', then the provability of a sentence from given axioms could no more provide a basis for it. As a consequence of the incompleteness result, truth and proof (within a mathematical axiom system) were no longer equivalent. Something stronger was required. The questions was effectively: what could provide us with a criterion of tautologousness that would still count as 'formal' but was independent of the notions of truth and falsehood? Carnap was convinced that a viable alternative could be found since he wrote that "in metalogic we cannot work with the concepts of truth and falsehood." [*ibid.*] How could such a definition be given? Some scholars have here misconceived the scope of the problem, taking it for granted that among mathematicians the notion of mathematical truth was unproblematically equated with provability in a formal system before Gödel's incompleteness results, thereby giving the impression that the problem Carnap was tackling was merely a problem of finding a new formal method of deciding the logical truth of a given sentence. Carnap was from the very start of his *Versuch* trying to devise a system in which the logical relations of sentences could be expressed purely formally without recourse to any interpretation. In mathematics, such a conception of formalism was evidently a minority view. Even if formal (formal in the more strict sense of the word) proofs constituted an essential part of the work of some eminent mathematicians (cf. Hilbert's work), they were always used in a context which involved interpretation of the symbols involved. Thus, the transition of Carnap marked by the assimilation of Gödel's results was not so much a matter of finding a strong enough axiom system for a certain fragment of mathematics to establish provability of all the relevant theorems, but a matter of dispensing with truth altogether, *contrary* to the attempts of most mathematicians.

Steve Awodey has presented an analogy between the problem situation of Carnap and that of a chess game. If we think of the starting position of the pieces (any position within a game that has been reached in accordance with the rules) on a chess board as the axioms, and the rules for permitted moves of the pieces as the rules of inference, then a sequence of moves that ends up in a checkmate may be regarded as a proof of a theorem. The crucial observation here is that there exist configurations of the (given) pieces in the board that correspond to a checkmate position but that cannot be reached by the rules of permitted moves from the initial configuration of the pieces. In a certain sense, a checkmate position of this kind corresponds to an analytic sentence that has no formal proof.

Retaining the concept of provability in the form of rules of inference of a formal system, Carnap thought it could be possible to define the notion of analyticity that would be in accordance with the former concept of a 'provable sentence' and yet be of wider scope than "provable". The goal was to devise a notion that could play the role that was formerly conferred on the Wittgensteinian concept of tautology. The new concept of analyticity was envisaged as providing a notion of truth for the meta-language.

5.1.3 Carnap's concern with analyticity

The source of the difficulties that related to the definition of analyticity for mathematical language(s) lay clearly in the incompleteness results of Gödel: for what they irrevocably showed was that one could not rely on a representation of a system of language based on definite axioms and deduction rules, if the professed specification of such a representation was to yield *all* mathematical truths. Thus, driven by the need to formulate a stronger notion of what is constitutive of a language, Carnap was impelled in the direction of investigation that consisted of attempts to provide a definition of analyticity for mathematics (purely mathematical sentences of a language).

Following the circulation of Carnap's *Untersuchungen zur Allgemeinen Axiomatik* among colleagues (including Gödel, who made some remarks concerning it in the opening paragraphs of his dissertation), Carnap and Gödel held discussions from time to time over the next few years. These discussions focused on — what else? — questions in logic and the foundations of mathematics. Carnap was among the first to hear of the incompleteness discovery of Gödel, as the latter told of it to Carnap in a conversation on 26 August 1930. Moreover, as mentioned in the Section **4.5.3**, Carnap was among the audience at the Königsberg conference where the result was first publicly announced. After Carnap's move to Prague in the summer of 1931, the two men began a correspondence that concerned mainly Carnap's attempts at refining his conception of the theory of linguistic forms, especially the notion of analyticity that figured prominently in those attempts, finally culminating in his *Logische Syntax der Sprache*.

Correspondence with Gödel

The first letter that Carnap sent to Gödel on New Year's Day 1932, consisted of an inquiry whether Gödel would be interested in reading the first part of his manuscript titled "Metalogik" (the would-be *Syntax*-book). Gödel presumably replied with interest, for the second letter from Carnap to Gödel in February begins with the following words:

> Dear Mr. Gödel, [i]n the coming days Hempel will send you the first part of my "Metalogic", which is what I have written so far. Feel free to keep the manuscript for a few weeks. When you have finished reading it, please do not send it back, but rather let me know. I would be very grateful for your critical comments. [1. Carnap to Gödel [Gödel, *Collected Works*, vol. IV, 343]]

In his reply, Gödel does not mention Carnap's manuscript, but requests that Carnap send him the second part, in case it is finished. The second part of the manuscript was referred to as "Semantics" by Carnap, and Gödel's letter in September 1932 to Carnap contains several interesting remarks on it. In this second part Carnap had attempted to provide a truth-definition for mathematics (i.e. for the purely mathematical sentences of a language that incorporated the theory of types), which amounted precisely to a definition of analyticity in Carnap's terminology. The definition proceeded by induction on the complexity of formulas. Gödel's remarks are worth quoting:

> First of all, I think that the definition of "analytic" for the extended language, and therefore the W-proof as well, is faulty. You specify rules by means of which the question of whether a formula is analytic is traced back to the same question for other formulas. Accordingly, then, the concept "analytic" would only be defined if this procedure were always to lead finally to formulas for which it is settled by other means whether they are analytic (for example, $0 = 0$). But it seems to me that this is not always the case. For let us take, e.g., the formula $(F)F(0) \vee \overline{F(0)}$; in order to establish whether it is analytic, one must do that for all formulas of the form $P(0) \vee \overline{P(0)}$. Towards that end, one must replace each constant predicate P by its definiens; but in the latter, bound predicate variables could again occur, and so on, so that one runs into an infinite regress. This becomes most evident in that, under certain circumstances, *the same* formula can always recur. [...] In my judgement, this error may only be avoided by regarding the domain of the function variables not as the predicates of a definite language, but rather as all sets and relations whatever. (*[A footnote inserted by Gödel:] This doesn't necessarily involve a Platonistic standpoint, for I assert only that this definition (for "analytic") be carried out within a definite language in which one already has the concepts "set" and "relation".) [3. Gödel to Carnap [Gödel, *Collected Works*, vol. IV, 347]]

Carnap had thus attempted to define the notion "analytic sentence" inductively, having recourse to a method that we nowadays refer to as a substitutional treatment of quantification. To illustrate the possible form that this initial definition of Carnap took, (nothing of the original first draft of the *Syntax* manuscript, apart from a table of contents, has survived[5]) consider an arithmetical sentence of the form $(\forall x)f(x)$, where quantification is taken over the numerical variable x and $f(x)$ is a formula in which at most the appearance of x is free. Then it would be quite reasonable to define:

$$(\forall x)f(x) \text{ is analytic} \iff f(a) \text{ is analytic for all numerical constants } a.$$

It is plausible, on the basis of the comments made by Gödel, that Carnap had attempted to generalize this idea further, providing a definition involving higher-order quantifiers that range over all properties or sets. In short, he considered also the analyticity of formulas of the form $(\forall X)f(X)$. Considering the singular case of $f(X) = X(0)$, one would have the definition

$$(\forall X)X(0) \text{ is analytic} \iff A(0) \text{ is analytic for all predicate constants } A.$$

[5]This, at least, is the well-founded assessment of Awodey and Carus [Awodey & Carus (2007)].

The problem with this definition is that it places no restrictions on the range of possible predicate constants $A(x)$ to be substituted for X. Thus, testing for analyticity, one could even consider substitution instances in which $A(x)$ gets the valuation $(\forall X)X(x)$. This makes the definition circular, and thus incapable of as serving as an adequate definition of analyticity. The so-called impredicativity of the higher-order quantifier is at the root of the problem here, so that in order to circumvent the problem, one should restrict the range of predicate constants by, e.g., limiting the allowable substitutions to predicates of lower "order", to be defined in a suitable sense. As Awodey and Carus aptly remark, such a procedure "would result in a workable scheme, but it would only provide a definition for a system like ramified type theory, which is inadequate for classical mathematics." [Awodey & Carus (2003), 17]

Carnap's first attempt to reconstruct Gödel's proposal is markedly desperate, as is evident from his reply to Gödel of 25 September. Carnap finds it difficult to come up with a suitable interpretation of the predicates in the object language L, in effect a difficulty residing in the professed necessity to capture "all values" of these predicates in L. It is not easy to make it explicit how this idea would be expressed, even with respect to another language L' in which the values are to be taken. Moreover, a definition with respect to a language L' would not be sufficient, for then predicates *not* definable would be left out, whereas one apparently needs all 'arbitrary' ones. The latter notion, however, is something "*ziemlich bedenklich*", according to Carnap. In the end, he is forced to ask for Gödel's help in finding the right definition, as everything else in the professed book will depend on it:

> Prag XVII, 25 September 1932
> N. Motol, Pod Homolkou
>
> Dear Mr. Gödel,
>
> Thanks very much for your letter. I am very grateful to you for having made me aware of the faultiness of the definition of "analytic". In the meantime I have attempted to work out a better definition, somewhat in the direction of your suggestions (if I have correctly understood them). But in this certain difficulties arise that I still cannot overcome. In any case, I realize that one will not arrive at the formulas of classical mathematics about real numbers if the universal quantifier with predicate variables (or function variables) ranges only over the predicates definable in a definite delimited system. You say: it must range over "all sets"; but what does that mean? My attempt goes as follows: by a predicate valuation I understand a rule that assigns to each numerical expression (for example, "$0'''$"), or, respectively, to each n-tuple of such, either "$0 = 0$" or "$0 \neq 0$"; a function valuation is a rule that assigns a numerical expression to each n-tuple of numerical expressions. With the aid of this concept, "evaluation" is defined: in the formula under consideration the assigned formula is put in place of "$F(\ldots)$" and the assigned expression in place of "$f(\ldots)$". "$[F](\ldots)$" is analytic if "$\ldots$" is analytic with respect to each valuation of F. In its entirety the definition is rather complicated, but it can be

done. But now the objection. If the definition is to achieve what was striven for, then for the rule — which must of course be formulated in the language of semantics — one may take not a restricted semantic language, rather the rule may be constructed with arbitrary semantic notions. (For otherwise certain sets of numbers would again remain outside, sets that are indeed plausible, but that cannot be comprehended within the system.) But is that not questionable? It seems to me not questionable if "analytic in the language S" cannot be defined in a semantics that is formalised in S, but only in a semantics that is formalised in a more extended language S_2. But to operate with a concept for which there is no language at all in which it can be rigorously defined is certainly rather questionable.

But perhaps you see a way to define the concept in a definite language. Or how is your remark to be understood: "...that this definition for 'analytic' may be carried out in a definite language in which one already has the concepts 'set' and 'relation' "? Can you define the concept "set" within a definite formalized semantics?

A further difficulty concerns the higher-type predicates; such things must be represented by assignments to valuations of lower type (or by a semantic set of sets, and so on, of expressions).

Will you perhaps be soon writing up a first draft of your paper, so that I could read it and make use of it? That would be extraordinarily valuable to me. Or can you at least tell me now by letter further details about the definition of "set" and "analytic"? For the purposes of my book, the concept "analytic" is very important, because I want to tie the usual concepts (especially "content", and so on) to it. In case your definition is usable for my purposes (which is surely to be expected), it would be most practical for me to adopt it from you (of course, with mention of your authorship), instead of first seeking one myself, which subsequently might prove to be worse than yours. In any case it would be of utmost value to me to see in what sense your definition really leads to a formalizable concept.

As to terminology: The term "true" seems to me very unsuitable; in any case, its usage would not be in accord with general linguistic usage. For according to the latter, the sentence "Vienna has so and so many inhabitants" is of course true, whereas the definition proposed by you surely does not apply to it. Thus one would surely have to say "logically true" or "logically valid" or "tautological" or "analytic"; and of those expressions the last seems to me the most suitable. [4. Carnap to Gödel [Gödel, *Collected Works*, vol. IV, 349–353]]

Gödel had hardly any time to ponder over his answer, for Carnap wrote him again, two days after the first reply, announcing that he had found a solution to the problem. Indeed, he had realized that the recalcitrant question about the interpretation of "all values" of a predicate could be answered in the context of a formal meta-language L' simply by using universal quantification over the predicates: $(\forall X)\ldots X\ldots$.

> Prag XVII, 27 September 1932
> N. Motol, Pod Homolkou
>
> Dear Mr. Gödel,
>
> The day before yesterday I wrote to you about a difficulty in the definition for "analytic". Yesterday I found the solution: The locution "for every valuation ..." that occurs in the definition can still be expressed in a semantics formulated in a definite language, namely by "$[F](\ldots)$", since a valuation is of course a semantic predicate. This is possible even though in the semantics under consideration not all possible valuations, that is, predicates, can be defined. To be sure, in application to predicates and functions of higher types and mixed types the matter becomes quite complicated indeed.
>
> [...]
>
> One more question: If in the language one allows predicate- and function-variables in operators, can one do without recursive definitions, without loss to definable predicates and functions? [5. Carnap to Gödel [Gödel, *Collected Works*, vol. IV, 355]]

The key new idea of Carnap here is that the language L' in which the valuations are taken needs to be stronger (in terms of expressive power) than the one for which they are given. [*ibid.*, 18] This seems to have been precisely the idea that Gödel had hinted at in his earlier letter, as he confirms Carnap's suggestion in a reply (considerably delayed) written on 28th November:

> As I gather from your second letter, you have understood my suggestions about the definition of "analytic" entirely as I meant them. In order to be able to carry out the matter in general, that is, for functions of arbitrary finite type, one needs a variable of the next higher type (type ω), i.e., one which runs through *all* finite types. This could be seen a priori, since one can never define "analytic" in the same system — otherwise contradictions will result. I believe moreover that the interest of this definition does not lie in a clarification of the concept "analytic", since one employs in it the concepts "arbitrary sets", etc., which are just as problematic. Rather I formulate it only for the following reason: with its help one can show that undecidable sentences become decidable in systems which ascend farther in the sequence of types.
>
> [...]
>
> The question you pose at the end of your letter of 27 September is to be answered affirmatively if the axiom of reducibility or an equivalent (e.g., your definition rule for functions) is contained within the system. [6. Gödel to Carnap [Gödel, *Collected Works*, vol. IV, 355–357]]

Gödel also refers in the letter to a "similar definition for 'analyticity' "given by Tarski in a paper that was to appear [Tarski (1932)], about which the latter had reported earlier in the

Anzeiger der Wiener Akadmie no. 2. Tarski's famous 'impossibility result' concerning the definition of the notion of truth in the object language was clearly one of the reasons why Gödel never worked out and published the professed sequel to his famous [Gödel (1931)]. The upshot of the correspondence was, then, something very different from what one would expect from the viewpoint of the Wittgensteinian universalism that had been very prominent in Carnap's thinking from 1929 on. As a matter of fact, the resulting definition of "analyticity" that at the outset had formed such a crucial element of the architectonic of the professed *Metalogik*, was now omitted from the first published version of the *Syntax* book because of "reason of space". The true reason underlying this decision was the recognition, hinted at by Gödel in his footnote on Platonism, that the notion of analyticity thus defined was not 'absolute', but rather, in a certain sense, conventional (i.e., relative to a language). This conception of analyticity, in as much as it undermined the universalism in Carnap's thinking, was perfectly in harmony with the conventionalism of his earlier work, including *Der Raum* and the *Aufbau*.[6] For now, the definition gave a notion of "analytic in L", but only with respect to *another* language L', i.e., a *metalanguage* used for the interpretation of L. This entailed that the question of *correctness* of a linguistic framework lost its all-encompassing significance; indeed, the issue of whether the choice of a language L' would be most natural in the sense that it is conducive to an illuminating notion of analyticity, could surely be raised, but it could hardly be claimed that such a choice would provide us with the *correct* notion of analyticity for a given language. This realization marked a watershed in Carnap's thinking, as it turned out to be of a far greater significance for the subsequent development of his philosophy than the specific metalogical results from which it sprang.

5.1.4 The principle of tolerance

The first public signal of the radical turn that Carnap had taken was the discussion contribution "Über Protokollsätze", published in the *Erkenntnis* No. **3**, a response to Neurath's "Protokollsätze" (in the same issue of *Erkenntnis*). Neurath's paper was, in turn, a response to Carnap's "Die physikalische Sprache", where Carnap had presented a proposal for compatibility between the physical and phenomenal languages, arguing for their usefulness for different purposes. In his criticism of Carnap's paper, Neurath had raised the question about *which* single language should one then have recourse to, granted that the single-language strategy was viable. Although Neurath had done much to inspire Carnap's turn to physicalism during the years 1930–1931, he was especially unhappy with the paper on physical language for he saw traces of foundationalism in it. Furthermore, he thought that he was unfairly anticipated as well as misrepresented in it. [Carus (2007), 252] After an emotional correspondence with Carnap in the early 1932, he sent a draft version of his reply in May. Carnap, as the co-editor of *Erkenntnis* set to work on drastic revisions, and the result was "Protokollsätze".[7] Carnap, in turn, began to write his reply to

[6]That even the *Aufbau* still contained elements of conventionalism is clearly visible from the observation, made in the book, that there are different possible bases for the construction of conceptual content.

[7]This paper contained "the proposal for a physicalistic, observer-indexed form of basic observation sentence, to counteract the danger of absolutising or hypostatising evidence reports", as succinctly put by Carus.

Neurath's critique and a defense of his position in the "Physikalische Sprache". The early draft of this reply seems to have survived as section I of the article "Über Protokollsätze". [Carnap (1932d)] The ripening of the core ideas regarding protocol sentences and physicalism seems to have taken place in the Tyrolean Alps where Carnap spent the later part of the summer in 1932. Incidentally, in September Herbert Feigl stopped by for several days with Karl Popper in tow, and the three had long conversations about protocol sentences during their extended walks in the mountains. During these conversations Popper put forward his suggestion about the relative status of basic statements. Thus, in keeping with his later repulsion for the purely linguistic aspects of philosophical analysis, Popper's account was probably already then framed in the manner of his presentation in the §§29–30 of *Logik der Forschung*. The relevant passage reads:

> Science does not rest upon solid bedrock. The bold structure of its theories rises, as it were, above a swamp. It is like a building erected on piles. The piles are driven down from above into the swamp, but not down to any natural or 'given' base; and if we stop driving the piles deeper, it is not because we have reached firm ground. We simply stop when we are satisfied that the piles are firm enough to carry the structure, at least for the time being. [Popper (1935), §30, 75–76; PT 111]

This account fits very well with the conception presented in the "Über Protokollsätze"; in fact there was a section in the paper that explicitly discussed Popper's position. The general contention of that paper was that physicalistic protocol sentences have precedence over other language forms but this was stated in a manner that did not exclude in a categorical way the possibility of having recourse to other language forms, including the phenomenalist one, as a last resort. What especially strikes one in the article is a completely novel tone with which the issues are discussed. This new tone in his writing would become deeply characteristic of all his subsequent work. For example, as a reply to the challenge presented by Neurath about the choice of a 'correct' language form, Carnap says that: "In my view the issue here is not between two conceptions that contradict each other, but rather between *two methods for constructing the language of science, which are both possible and justified*." [Carnap (1932d), 215][8] The grounds of this pluralism, brought forward by the discussion on analyticity with Gödel, as we saw above, were clearly spelled out:

> Not only the question whether the protocol sentences are inside or outside the syntax language, but also the further question regarding their precise specification, is to be answered, it seems to me, not by an assertion, but by a stipulation [*Festsetzung*]. Though I earlier [in "Die physikalische Sprache"] left this question open, [...] I now think that the different answers are not contradictory. They are to be taken as proposals for stipulations [*Vorschläge zu*

[*ibid.*]

[8] "Nach meiner Meinung handelt es sich hier aber nicht um zwei einander wiedersprechende Auffassungen, sondern um *zwei erschiedene Methoden zum Aufbau der Wissenschaftssprache, die beide möglich und berechtigt sind.*"

Festsetzungen*]; the task is to investigate these different possible stipulations as to their consequences and assess their usefulness. [*ibid.*, 216][9]

The tendency towards assimilating elements of pragmatism in his philosophy is here clearly visible. The break with the view of the pre-1931 programme of Vienna Circle is decisive: the essential content of his original reply to Neurath is retained but now expressed more pronouncedly in a pragmatic context. The gist of this move comprises the idea of evaluating different *proposals* for a system of language in a given domain; these evaluations, in turn, are coördinated pragmatically; whether *this* or *that* language form suits better the purposes and goals of a particular situation. Carnap is very much aware of the radical departure from his and the Vienna Circle's earlier position, as he states in a further paragraph:

> In all theories of knowledge to date there is a certain *absolutism*: in the realistic theories an absolutism of objects, in the idealistic ones (including phenomenology) an absolutism of the "given", of "experiences", of "immediate phenomena [*unmittelbare Phänomene*]". Even in positivism we find this residual idealistic absolutism; in the logical positivism of our circle — in the works of the logic of science (epistemology) published to date by Wittgenstein, Schlick, Carnap — it takes the more subtle form of an absolutism of primitive propositions ("elementary propositions", "atomic propositions"). [*ibid.*, 228][10]

Here is, then, expressed in its essentials, the view that would later be termed the "Principle of Tolerance" by Carnap. In the *Logische Syntax der Sprache* it would be explicitly put forward as a principle for coördinating the discourse on the foundations of mathematics, where the controversy between the three principal 'philosophies of mathematics' — formalism, intuitionism and logicism — had resulted in an embarrassing muddle of conceptions. Indeed, as Carnap stated in the preface to the book, progress in logic had been severely hampered by the timid notion that logic had to be 'correct' and that the object was to find the 'true' logic. The goal of the *Syntax* book is to untie the ship of logic from its mooring to the prejudice of 'correctness' by eliminating this notion. Overcoming this

[9]"Nicht nur die Frage, ob die Protolkollsätze außerhalb oder innerhalb der Systemsprache stehen, sondern auch die Weitere Frage nach ihrer genaueren Kennzeichnung ist, wie mir scheint, nicht durch eine Behauptung, sondern durch einen Festsetzung zu beantworten. Während ich früher [...] diese Frage offen ließ und nur einige mögliche Antworten andeutete, meine ich jetzt, daß die der verschiedenen Antworten einander nicht widersprechen. Sie sind als Vorschläge zu Festsetzungen aufzufassen; die Aufgabe besteht darin, diese verschiedene möglichen Festsetzungen auf ihre Folgerungen hin zu untersuchen und ihre Zweckmäßigkeit zu prüfen."

[10]"In allen bisherigen Erkenntnistheorien steckt ein bestimmter *Absolutismus*: in den realistischen ein Absolutismus der Objekete, in den idealistischen [einschließlich der Phänomenologie] ein Absolutismus des "Gegebenen", der "Erlebnisse", der "unmittelbare Phänomene". Auch im Positivismus findet sich ein Rest dieses idealistischen Absolutismus; in den logistischen Positivismus unseres Kreises — in der bisher veröffentlichten wissenschaftslogiken (erkenntnistheoretischen) Schriften von Wittgenstein, Schlick, Carnap — nimmt er die verfeinerte Form eines Absolutismus der Ursätze ("Elementarsätze", "Atomsätze") an."

timid conception, we set sail and "before us", states Carnap with palpable enthusiasm, "lies the open sea of free possibilities."

This then, was the context in which the Principle of Tolerance was conceived. As we have seen, the correspondence with Gödel very much prepared Carnap to take the step toward linguistic pluralism. This pluralism is then the background framework for the studies of philosophy of mathematics undertaken in the *Syntax* book. As far as the foundations of mathematics are concerned, the situation could be described as follows. In the light of the incompleteness results of Gödel, the hopes of constructing a universal language for the purposes of promoting the program of unified science abraded. Moreover, the status of the program of logicism altered dramatically. Carnap was, as already pointed out, among the first philosophers to fully realize this and its implications for any epistemological project attempting to set our knowledge on a firm ground of logic. Even the nature of mathematics and its epistemological basis became once again, after a period of unflinching confidence in logicism, an issue of concern, although from a completely new angle. However, what has sometimes been seen as a negative impact of Gödel's results with regard to theoretical pursuits in the various sub-domains of mathematics, has in fact turned out to be a liberating insight about the possibilities and aims of a mathematician. Indeed, as soon as it is realized that the incompleteness results do not concern the *descriptive completeness* of a language that a mathematician uses to characterize the objects and structures that he is interested in, but 'only' the *deductive completeness* of that very same language, it becomes clear that a large part of what can be said to comprise the most important and most interesting part of mathematical theory construction, is retained. In keeping with this insight, Norbert Wiener, a prominent mathematician and a prodigy, known for his seminal work in the measure-theoretical investigations in probability theory, once characterized the nature of mathematical work in the following beautiful way:

> To other mathematicians, the task of the mathematician is to use a rigid and a demanding medium to express a new and significant vision of some aspect of the universe; to express *aperçus* which reveals something new and something exciting. If this medium is strict and confining, so are in fact the media of all creative artists. [Wiener (1956), 62]

The gist of Wiener's description lies in the observation that the creative dimension of mathematics is not in the least affected by the 'negative' results of Gödel. On the contrary, what seems to be the contention of many great mathematicians is that Gödel's results confirm their belief based on their working experience that they are really investigating a "primordial mathematical reality" with the instruments of apprehension that the different axiomatic representations provide. A prominent example of this basic attitude is Alain Connes' remark, expressed in a discussion with André Lichnerowicz and Marcel Paul Schützenberger, about the import of Gödel's theorem:

> The content of Gödel's theorem is not only undecidability, that is, the existence of undecidable propositions, whatever the system of axioms one uses.

> Far more important in any logico-deductive system (one that is sufficiently elaborate to be able to handle the natural numbers) is the existence of propositions that are true but that cannot be proved within the system. The mental image of this statement is that of a world that I call "the primordial (or primitive) mathematical world" and which is never exhausted by the instruments of comprehension one constructs, based on logical reasoning and on a given axiomatic system. Indeed any such system is always found wanting since Gödel's theorem guarantees the existence of an infinite number of true propositions dealing directly with arithmetic, but which are not provable within the logical system. [Connes et al. (2001), 6]

As distant as Connes' basic attitude seems to be from that of Carnap who eschewed all references to a "primordial reality", there is one particular aspect of Connes' account that is strikingly similar to the basic *ethos* of Carnap's linguistic pluralism. This similarity is exemplified in Connes' conception of axiomatic systems as different linguistic tools for grasping the mathematical reality. In a similar manner, Carnap, while bracketing any talk of an "objective reality", is willing to assess the usefulness of different linguistic presentations for different practical tasks. Notwithstanding the difference of their motives, the application of different frameworks is a hallmark of the scientific activities of *both* the Carnapian and the Platonist mathematician:

> It is worth noting that most mainstream mathematics only requires the countable axiom of choice. Solovay has constructed an axiomatics in which all the sets are measurable.[11] This axiomatics is just as non-contradictory as that of set theory with the global axiom of choice. It does not depend on a hypothesis about inaccessible cardinals, but I will spare you the details. [...]
>
> This is another illustration of my opposition between primordial mathematical reality and the axiomatic system that enables us to perceive it. It is clear that our perception of this reality will change depending on the lens we use. If we take an axiomatic system with the uncountable axiom of choice, we will have a certain view, a certain grasp of this primordial reality, and if on the other hand we take Solovay's axioms with the countable axiom of choice and measurability, we will have another. But there is absolutely no contradiction.
>
> The primordial reality is unaffected; it remains immutable. It remains identical to itself. [*ibid.*, 21]

5.1.5 *Logische Syntax der Sprache*

The sense of breakthrough that characterized the enthusiastic expressions in "Über Protokollsätze" [Carnap (1932d)] is still equally present in the famous introduction to the *Logische Syntax der Sprache*, where the new insights brought forward in connection with the protocol-sentence debate are reiterated with breaching confidence:

[11]The reference is to J. Stern: "Le problème de la mesure", *Séminairé Bourbaki 1983/1984*, Exp. 632.

> Der Kreis möglichen Sprachformen und damit der verschiedenen möglichen Logiksysteme ist [. . .] unvergleichlich größer als der sehr enge Kreis, in dem man sich in den bisherigen Untersuchungen der modernen Logik bewegt hat. Bisher ist man von der schon klasscisch gewordenen Sprachform, die *Russel* gegeben hat, nur hin und wieder in einigen Punkten abgewichen. [. . .] Der Grund dafür, daß man sich bisher nicht weiter von der klassischen Form zu entfernen wagt, liegt wohl in der weit verbreiten Auffassung, man müsse die Abwiechungen "rechtfertigen", d.h. nachweisen, daß die neu Sprachform "richtig" sei, die "wahre Logik" wiedergebe. Die Auffassung und die aus ihr entspringenden Scheinfragen und müßigen Streitigkeiten auszuschalten, ist eine der Hauptaufgaben dieses Buches. [Carnap (1934a), v]

The programmatic tone of the introduction adds to the impression that a radical step has been taken with respect to traditional accounts of language. As Carnap says, the first attempts to escape from the "classical" language forms were certainly daring. But they were still hampered by the prejudice of 'correctness': "Aber sie waren gehemmt durch das Streben nach 'Richitigkeit'." The introduction is concluded with the picturesque words: "Nun aber ist die Hemmung überwunden; vor uns liegt der offene Ozean der freien Möglichkeiten." The explicit expression of the Principle of Tolerance is found in the §§16–17 where it is accompanied with a general discussion about the necessity to state the meta-theoretic or *wissenschaftslogische* proposals in precise terms. These proposals have to be expressed as explicit rules or definitions, *within* the formation or transformation rules of a precisely defined language or calculus. As a number of scholars have observed, the constructive requirements represented in these paragraphs exemplify the *voluntaristic* dimension of Carnap's thought:

> Hat man sich einmal klar gemacht, daß alle pro- und kontra-intuitionistischen Erörterungen von der Form eines Kalküls handeln, so wird man die Frage nicht mehr in der Form stellen "Wie *ist* das und das?" sondern "Wie *wollen* wir das in der aufzubauenden Sprache einrichten?" [. . .] Damit verschwindet die dogmatische Einstellung, durch die Diskussion häufig unfruchtbar wird. [Carnap (1934a), 42][12]

One of the main motives of the *Syntax*, as we have already hinted at above, was to bring order into the chaos that prevailed in the field of philosophy of mathematics. What Carnap especially wanted to accomplish was to set up a framework in which to express clearly and unambiguously the constructivist viewpoint in mathematics represented by Brouwer, Kaufmann, Wittgenstein and others. This line of work derived from the investigations of Leopold Kronecker (1823–1891). The constructive arguments were further cultivated by French mathematicians, most prominently by Poincaré and Borel. In 1912

[12]"Once it is understood that all pro- and anti-intuitionist considerations are concerned with the form of a calculus, the question will no longer be asked in the form 'What *is* the case?' but rather 'How do we *want* to set this up in the language being constructed?' [. . .] And with that, the dogmatic frame of mind that often makes the discussion unfruitful is banished."

Brouwer was still inspired by the classical conception of constructivism *à la* Poincaré which was based on the *desideratum* of prescribing a law whenever a set with infinitely many points is worked with. A classical example is the determination of the set of decimals of the irrational number $\sqrt{2}$ which is infinite but determinate because there is a specific procedure to calculate all the decimals. In 1916 Brouwer presented a version of constructivism based on the *principium tertii exclusi*. (The presentation was given as a part of a course on the theory of point sets which was run during the academic year 1915–1916.) His notes dating from that time express clearly the new point of view:

> A mathematical thing is either an element of a previously constructed fundamental sequence $\mathcal{F}$ (governed by induction, like the sequence ρ) or a fundamental sequence f (which is not finished and not governed by induction) of arbitrarily chosen elements from $\mathcal{F}$ or a finite set. With such a sequence one can work very well, if one always has *at each phase* for the *finite thing d or the fundamental sequence r* that is derived from it, to work with a suitable initial segment of f (r is then in general *also* never finished) [...] A set is now a law by means of which a d or an r is derived from f; this r can then, for example, contain also relation symbols (e.g. ordering ones), so that the law can, for example, lead to well-ordered sets or other ordered sets, or to functions (one can indeed not get the set *of* ordered sets or the set of well-ordered sets). In addition one can accept *pseudo-sets* defined by comprehension, better called *species*, and one can call one species of a higher cardinality than another one, or two species equivalent. [Cited in van Dalen (1999), 240–241]

Carnap's Language I in the *Syntax* was effectively a framework for expressing more rigorously the philosophical tenets of constructivism as advocated by Brouwer and others. The success of the articulation of the intuitionist tenets and their implications in this framework was however questionable, as Carnap himself pointed out, because there was no way of telling whether the reformulation of them as precise definitions and rules really captured the views of intuitionists. The problem was that many of the articulations of philosophical views by mathematicians were made only in terms of vague "Erörterungen" that leave many crucial technical questions open "when one gets down to the brass tacks of constructing an actual language". [*ibid.*, 44] These tendencies were then clearly addressed in Carnap's general statement of the Principle of Tolerance:

> Unsere Einstellung zu Forderungen dieser Art sei allgemein formuliert durch das *Toleranzprinzip: wir wollen nicht Verbote aufstellen, sondern Festsetzungen treffen. [...] In der Logik gibt es keine Moral.* Jeder mag seine Logik, d.h. seine Sprachform, aufbauen wie er will. Nur muß er, wenn er mit uns diskutieren will, deutlich angeben, wie er es machen will, syntaktische Bestimmungen geben anstatt philosophischer Erörterungen. [Carnap (1934a), 45][13]

[13]"Our attitude to demands of this kind may be stated generally by the *principle of tolerance: we do not want to impose restrictions but to state conventions. [...] In logic there are no morals.* Everyone can construct his logic, i.e., his language form, however he wants. If he wants to discuss it with us, though, he will have

This formulation makes it evident that a crucial step is already taken towards the successor project of "rational reconstruction", viz. the project of "explication". The general *desideratum* of *replacing* a vague concept (or even a system of concepts) with a precise equivalent is, naturally, nothing new, as the earlier Vienna Circle view had accommodated precisely this requirement in its program of "rational reconstruction". Rather, what is new is the exhortation to give precise rules within a linguistic framework to be able to *judge the practical merits or drawbacks of a proposal*. Indeed, as the rigid conception of 'correctness' is dropped once and for all, the only criterion that remains for the task of selecting among different frameworks is practical usefulness. The pragmatic dimension that became to characterize the successor program of "explication" was not formulated explicitly by Carnap until 1945.[14] Why did it take Carnap so long to arrive at the notion of *explication*? In the *Syntax*-book the ideal of explication, sketched in the statement of the Principle of Tolerance, took a form that was soon recognized by Carnap to be inadequate. Its implementation was called in the *Syntax* "translation from the material mode into the formal mode of speech". This notion of explication was still very much influenced by certain residual ideas from the earlier period. Indeed, the original syntax idea that Carnap conceived in the sleepless night in 1931 represented, above all, a rejection of meaning in Wittgenstein's sense, i.e., the thesis about the *ineffability of semantics*.[15] One of the clear indications that Carnap indeed advocated a view that could be termed as the thesis of the ineffability of semantics, is the fact that in the heyday of the Vienna Circle in the early thirties, already before the publication of the *Syntax*, he preferred what the members of the circle called the *formal mode of speech* to what they called the *material mode of speech*. In general, this marked an inclination to express one's theoretical views on language in syntactical rather than semantical terms. A often-quoted example is Carnap's preference for the formal mode of saying "'Five' is a number word" instead of the material mode of saying that "Five is a number". This simple example illustrates how the entire program of the "logical syntax of language" belonged to the orbit of ideas associated with the ineffability of semantics and the idea of language as a universal medium. This fundamental idea allegedly has its roots in Wittgenstein's *Tractatus*. As is well known, in an angry letter to Moritz Schlick, dated 8 August 1932, Wittgenstein accused Carnap of borrowing the idea of the formal mode of speech from his *Tractatus* without due acknowledgment:

> You know yourself very well that Carnap is not taking a single step beyond me when he approves of the formal and rejects the "material mode of speech". It is inconceivable to me that Carnap should have misunderstood the last few propositions of the *Tractatus* — and hence the basic idea of the entire work — so thoroughly [as not to know it, too]. [Nedo & Ranchetti (1983), 254–255]

Was Wittgenstein's reaction justified? Wittgenstein's notorious self-centeredness notwithstanding, it is arguably true that a significant part of Carnap's conception of the impor-

to make precise how he wants to set things up. He has to give syntactic rules rather than philosophical considerations." [PT, 51–52]

[14] That is, in the article "Two Concepts of Probability" [Carnap (1945)]. The most detailed exposition of the new method of explication was provided in as late as 1950 in Chapter 1 of *Logical Foundations of Probability*. I will survey the essentials of that exposition in the Chapter 6.

[15] Carnap later referred to this view as "absolutist". [Awodey & Carus (2003), 22]

tance of syntactic methods in the theoretical study of language derived from the *Tractatus*. But what Wittgenstein missed was another important ingredient in Carnap's view. Carnap was not so much opposed to the material mode of speech as he preferred the formal mode of speech. Regarding his contact with Gödel and the impact that the latter had on his thinking, it is not extraneously speculative to conjecture that Gödel's method of arithmetization that enabled him to construct a syntax for the language of (Peano) arithmetic in that language itself, was a major source of inspiration in Carnap's attempts to construct a general theory of linguistic forms on a similar basis. Carnap's ultimate aim was, after all, to construct a truly universal language for language theory. Thus, he was not so much concerned with the fundamental limitations of the universalist approach to linguistic analysis as with the constructive possibilities that such a line of attack made available. As is well known, this underlying idea is exemplified in the *Language I* of the *Logische Syntax*, which briefly stated, includes, on the mathematical side, the elementary arithmetic of the natural numbers with certain restrictions imposed on the sentences that are expressible in that language. The most important restriction consists of the condition that only definite number-properties occur. This refers to such properties of numbers the possession or non-possession of which can be determined in a finite number of steps according to fixed rules given within the language. It is useful to note that this restriction does not, however amount to a violation of Gödel's incompleteness results. Indeed, the language is indefinite in the sense that it contains sentences that are not resoluble (i.e. either demonstrable or refutable). As to the illegibility of the material mode of speech, Carnap gave an explicit rendering of his stance in §81 of the *Syntax* (titled "The Admissibility of the Material Mode of Speech"): "We have spoken of dangers and not of errors of the material mode of speech. *The material mode of speech is not in itself erroneous*; it only readily lends itself to wrong use. But if suitable definitions and rules for the material mode of speech are laid down and systematically applied, no obscurities or contradictions arise." [Carnap (1937), 312] What is furthermore emphasized by Carnap in that very same paragraph is that especially when important conclusions or philosophical problems are represented by sentences in the material mode of speech, it is advisable to transform them into the formal mode of speech so as to make them free from ambiguity. Moreover, Carnap reiterates his point by saying that: "It is not by any means suggested that the material mode of speech should be entirely eliminated." [*ibid.*] To illustrate his thesis that "translatability into the formal mode of speech constitutes the touchstone for all philosophical sentences", consider the following example of how that method is used in practice: [Carnap (1937)]

Philosophical sentences (Material mode of speech)	*Syntactical sentences* (Formal mode of speech)
A. *Generalities* (about things, properties, facts, and so on) [. . .]	
27 a. A property of a thing-property is not itself a thing-property.	27 b. A $^{2}\mathfrak{pr}$ is not a $^{1}\mathfrak{pr}$.
28 a. A property cannot posses another property (As opposed to 27 a.)	28 b. There is no $\mathfrak{pr}$ of a level higher than the first. (As opposed to 27 b.)
29 a. The world is the totality of facts, not of things.	29 b. Science is a system of sentences, not of names.
30 a. A fact is a combination of objects (entities, things).	30 b. A sentence is a series of symbols.
31 a. If I know an object, then I also know all the possibilities of its occurrence in facts.	31 b. If the genus of a symbol is given, then all the possibilities of its occurrence in the sentences are also given.
32 a. Identity is not a relation between objects.	32 b. The symbol of identity is not a descriptive symbol.

This example shows clearly how Carnap had overcome the restriction of the 'picture-theoretic' viewpoint of Wittgenstein while retaining the latter's notion of language-relativity of knowledge. The liberation from the Tractarian straightjacket of meaning-constitution via truth-functional concatenation of atomic sentences representing atomic facts was made possible by extending a Hilbertian or Tarskian formalist view from logic and mathematics to the whole of knowledge. The question remains, however, about the extent to which the notion of tolerance propounded by Carnap depends on a specifically "syntactic" approach. Awodey and Carus maintain that "tolerance depends on two components of the original insight of the sleepless night: (a) the distinction between language (a calculus, a purely syntactic symbol system) and its interpretation; (b) the requirement that a language be entirely specified by explicit rules. These two components survive unscathed and undiminished into the semantic period. (So it is rather misleading to call them 'syntactic'; Carnap's original term 'metalogical' might be more appropriate.)" [Awodey & Carus (2003), 24] The original all-encompassing skepticism about the possibility of explicating "meaning" (inspired by Wittgenstein) does not survive the transition. Moreover, the restriction of interpretation to the object language which was Carnap's first response to the necessity of distinguishing between a language and its interpretation, was mitigated further, as he came to see that interpretation could be specified by explicit rules (governing the application of such semantical notions as *satisfaction*, *designation* and *truth*). [*ibid.*] Thus, the transition to an overtly semantical approach to philosophical problems in Carnap's subsequent work was not an abrupt change in the basic orientation of his philosophy. In keeping with the Principle of Tolerance, the reinstatement of an explicated form of meaning was a sign of the recognition that the informal pre-theoretical notion of meaning was not in itself the source of difficulty but rather a particular, inherited and singularly obscure conception of it. This new explication met the standards set for evaluating the practical utility of the notion of "meaning", i.e., the very standards on the basis of which the previous conception had been rejected. As Awodey & Carus eloquently put it, "The original rejection of 'meaning' had proscibed what seemed an occult property, just like the rejection by Lavoisier of the traditional explanation of burning as the *release* of a substance ('phlogiston' in Stahl's theory) into the surrounding air. [...] In the same way, the later instatement of the idea that burning (oxidation) involved the *release* of electrons by the substance being oxidized met the standards of the post-Lavoisier principle of the conservation of matter, by which all reactions are regarded as recombinations of indestructible atoms. The new explication of the informal concept of 'meaning' has no more in common with the previous occult property than electrons do with phlogiston." [*ibid.*, 25]

5.2 The Semantic Turn

5.2.1 *Introduction to Semantics* and the correspondence with Neurath

In 1942 Carnap was in a position to re-evaluate the philosophical import of his general principles underlying language construction, i.e. the *principle of general syntax* and the *principle of tolerance*. Instigated by the thoroughgoing discussions with, *inter alia*, Morris, Nagel, Neurath and Tarski, Carnap had come to appreciate the significance of semantic analysis in such language-engineering tasks. He was thus able to considerably relax the formal requirements pertaining to syntax that in large part still dictated the general outlook of the *Logical Syntax*. In *Logical Syntax* the principle of tolerance did not yet play a prominent role; indeed, it was a kind of late comer in Carnap's thought, not necessary to the overall conception of the book. But later, it was to turn over the whole basic approach of Carnap, ultimately winning out the basic conception of the formal mode of speech. In his *Introduction to Semantics* (1942), the necessary modifications of Carnap's earlier views that figured in the *Logical Syntax* are discussed briefly in a section at the end of the book. Especially, "*[t]he Principle of Tolerance* (perhaps better called 'principle of conventionality') [...] is still maintained. It states that the construction of a calculus and the choice of its particular features are a matter of convention." [Carnap (1942), 247] But, in view of the essential characteristics of the principle, an important qualification is made, to the effect that "[...] the construction of a system of logic, i.e. the definitions for the *L*-concepts, within a given semantical system is not a matter of mere convention; here the choice is essentially limited *if the concepts are to be adequate* [...]."[16] With regard to the formal mode of speech, more radical modifications are required: "[Syntax] *Part IV* gives an outline of *general syntax*, which 'is to be regarded as no more than a first attempt' [...] Here, as

[16] My italics.

was to be expected, *greater changes are necessary*. Some definitions (especially those for *L*-concepts) have to be abandoned. In general, the syntactical discussions remain valid; but in many cases, they should be supplemented by semantical discussions." [*ibid.*][17] The most important changes occurred with regard to the distinction between logical and descriptive signs, and the related distinction between *logical* and *factual truth*. Carnap came to see it as necessary to make these distinctions in semantical, not syntactical, terms. The reasons were quite clear for Carnap. Even if the relevant concepts allowed for a formalization, in effect a representation by syntactical concepts in an appropriate calculus, "the question whether or not a given syntactical concept, e.g. '*C*-true in *K*' is the formal representation of the corresponding *L*-concept, say '*L*-true in *S*', is a question which cannot be answered in syntax alone". [*ibid.*] On the level of general objectives of philosophy, the classical theses of *Logical Syntax* remained valid in *spirit* albeit not in form. The first thesis (a. "Theoretical *philosophy* is the logic of science") remained valid without qualification. For Carnap, it is only a question of terminology whether "philosophy" could be used in a wider sense, including certain problems of empirical origin. If so, then these problems would most naturally fall into the area of pragmatics. [*ibid.*, 250] The second thesis (b. "Logic of science is the syntax of the language of science.") had to be modified, however. Carnap suggests that "[*t*]*he whole thesis is changed to the following: the task of philosophy is semiotical analysis*." [*ibid.*]

It has already been observed by some researchers, notably by A. W Carus, that there is very little evidence, published, or unpublished, bearing on this important transition in Carnap's thought. Carus claims that the very transition was brought about by the adoption of the principle of tolerance, but that its significance had not been immediately clear to Carnap. There was a considerable period of latency between figuring this principle out in the first place, and its emergence as the Archimedean point of Carnap's overall program. This is undoubtedly true. However, the reasons Carus gives for the considerable delay in the full development of the principle in Carnap's thought, are controversial:

> Though Carnap had abandoned the idea that we cannot step outside language, this idea had in turn rested on a foundation that was not so easily dismantled. The sharp distinction between concepts and sentences internal to a system and those outside it had been one of the main lessons Carnap had learned from Frege, and the application of this Fregean distinction to epistemology had been the basis of the Vienna Circle's formulation of rational reconstruction. [...] It had made possible their characteristic approach to the replacement of folk concepts in the context of a *deductive system* of knowledge. Wittgensteinian considerations about the impossibility of stepping outside language had then reinforced that sharp distinction even further. But even when that further element had been dispensed with, it was difficult now to backtrack from the sharp Fregean distinction between internal and external, and make this dichotomy less absolute. It was hard to *retain* the idea of a sharp distinction between internal and external for constructed languages while allowing room for other systems (ordinary language) that *lacked* this

[17] Italics in the original.

> sharp distinction. And only this concession would make it possible, in the late 1930s, to regard ordinary language and formal systems as mobile points along a continuum rather than as fixed diametrical opposites. [Carus (2007) 263–264][18]

Although concise and consistent, such an explanation is not fully acceptable, however. There were at least two conspicuous routes by which Carnap could have come to the conception of the relative elasticity of the boundaries between internal and external questions with regard to language engineering. Moreover, it seems that he must have been at least dimly aware of this elasticity and its prominent role in language construction well before the late 1930s. The first route was provided by the long-time interest of Carnap in artificial languages, especially his concern with Esperanto.[19] The second was provided by his obsession with a critical enquiry concerning physicalism, an issue which Neurath enticed him to turn his attention to. Immersion in both of these fields of interest must have brought it clearly within his view that the questions pertaining to the pragmatic concerns of constructing a language appropriate to any particular task are not straightforwardly soluble within the formal mode of speech alone. Even if the variety of artificial languages and the issues relating to their pragmatics did not convince him of this, the discussions between him and Neurath about the fortunes of physicalism must have, at least unconsciously, prepared him for the new semantic conception. Furthermore, Wittgenstein had by then already abandoned the Tractarian program of language as a universal medium. Already in his *Blue* and *Brown Books* dictated between 1933–1935 we find very clear presentations of a conception of language as an interpreted calculus in connection with which the transitory boundaries between internal and external questions in Carnap's sense were clearly manifested. Indeed, a hallmark of Wittgenstein's analyses of the different "language-games" was the use of swift, back-and forth transitions between the expressions of a particular language form and its description in a colloquial language. Although Carnap may not have been fully aware of these developments in Wittgenstein's thought at the time of their inception, Carnap himself was very much occupied with similar considerations in his correspondence with Neurath. It must be admitted, however, that the tension between the diametrically opposite elements alluded to

[18] Italics in the original.

[19] There is an interesting excerpt in the Autobiography which includes reminiscences about Carnap's involvement with the international Esperanto movement. This excerpt makes it evident how Carnap's insistence on the possibility of devising artificial languages for completely practical purposes of life reflect the idea of the relatively elastic boundaries between artificial and natural languages: "After the World War, I had some opportunities of observing the practical use of Esperanto. The most extensive experience was in 1922, in connection with the Esperanto Congress in Helsingfors, Finland. There I became acquainted with a Bulgarian student; for four weeks we were almost constantly together and became close friends. After the congress we traveled and hiked through Finland and the new Baltic republics of Estonia, Latvia and Lithuania. We stayed with hospitable Esperantists and made contact with many people in these countries. We talked about all kinds of problems in public and in personal life, always, of course, in Esperanto. For us this language was not a system of rules but simply a living language. After experiences of this kind, I cannot take very seriously the arguments of those who assert that an international auxiliary language might be suitable for business affairs and perhaps for natural science, but could not possibly serve as an adequate means of communication in personal affairs, for discussion in the social sciences and the humanities, let alone for fiction or drama. I have found that most of those who make these assertions have no practical experience with such a language." [Carnap (1963), 69]

by Carus above, was certainly bleak and unrelenting, to the extent that Carnap was keen to hold the Fregean distinction as the more promising way to go. The point of transition is important, not only from the point of view of adopting the Principle of Tolerance, but also because it marks another, related transformation in Carnap's thought. This is the transition from the conception of language as a universal medium to the conception of language as a calculus, the terms already having figured above in our discussion about Carnap's connection with Wittgenstein.

An important incitement to the development of the semantical conception of Carnap towards the more general program of explication came from the discussions between him and Neurath over the basic semantical concepts of truth, meaning, translation etc. A most useful source of information in this respect is their correspondence, the most important and interesting part of which dates from the early 1940s, coinciding with the 'semantic turn' of Carnap.[20] A symptomatic expression of his liberated notion of language engineering is found in the letter to Neurath (November 7th, 1942), where a suggestion for an auxiliary language by Lancelot Hogben is discussed:

> I am very much interested in what you wrote about Hogben's ideas on language-making and about his own auxiliary language. I remember our previous discussions where I maintained the superiority of artificial languages in comparison to Basic English, while you were rather sceptical about their practical chances. You say that Hogben's *Interglossa* seems better than the other artificial languages; do you know enough of the other ones to make a critical comparison or do you merely infer it from Hogben's good ideas about the method of language-making? As Morris wrote you already, we think it better to see first Hogben's book before we decide on a monograph of his for the *Encyclopedia*. Who will be the American publisher of Hogben's book and when is it to appear? I am looking forward to it with great interest. [WKA (1942a)][21]

The monograph Carnap is alluding to, and which was under consideration to be included in the series *International Encyclopedia of Unifed Science*, was presumably *Principles of Animal Biology* (1940). It was published in the series in the volume dedicated to Biology.[22] The *Interglossa* did not appear until 1943. By October 7th, 1944 Carnap had already received the book from Neurath and read it, enabling him to make some critical comments:

> I read it with very great interest, and I own and read likewise "The Loom of Language" which I found likewise very interesting and which in some points gives no more detailed explanations of the reasons for Hogben's decisions. In

[20]This correspondence is kept up in the *Noord-Hollands Archief* forming part of the collection *Vienna Circle Papers* [References to this correspondence are henceforth abbreviated as WKA (= Wiener Kreis Archiv)].

[21]Cited with the compliance of the *Wiener Kreis Stichting* (Amsterdam). All rights reserved.

[22]As Carnap writes to Neurath in a letter dated February 4th, 1944: "I agree with the choice of Hogben for the biology monograph, as Morris will have written to you. When and where will his 'Interglossa' appear? I am looking forward to it with great interest." The *Interglossa* had already appeared in 1943, published by London: Penguin/Pelican Books.

> some points this new language has made important improvements in comparison with the earlier projects. I am not quite sure whether the choice of mostly greek word roots is the best possible. I have always thought that any new project should make improvements in two important respects: (1) to utilize the improvements made in Basic English, especially the principle of word economy; (2) certain improvements in the logical structure of the language which might be learned from symbolic logic. Hogben has done very well in the first point; and his construction of a very simple syntax is a great achievement. However, with respect to logic his language seems to me to have some weak points; e.g. the line between observable things and properties is not always drawn in the right place, and the whole matter of this distinction between tow word classes could probably be simplified still more; further his treatment of "all", "every", "any", "some" etc. is not satisfactory. The examples of translations which he gives at the end deviate in many points considerably from the original; thus a re-translation into English would lead to quite different texts. Tests of this kind made with earlier languages (e.g. Esperanto and Ido) had much better results. [WKA (1944x)][23]

Notwithstanding Hogben's ideas, which certainly plucked a string in Carnap's mind, the correspondence between him and Neurath concentrated very much on clarifying the points upon which the two men disagreed and on what grounds. The correspondence was always conducted in the most sincere and friendly manner despite the occasional outbursts (often committed tongue-in-cheek) of one or the other (more frequently Neurath) caused by the alleged absurdity of the position of the correspondent. An example of this is the letter of Neurath to Carnap on January 15th, 1943 (in answer to Carnap's dated November 7th, 1942):

> [...] I am just looking through the main chapters [of *Introduction to Semantics*], particularly the chapters you mentioned in your letter. I am really depressed to see here all the Aristotelean metaphysics in full glint and glamour, bewitching my dear friend Carnap through and through. As often, a formalist drapery and hangings seduce logically minded people, as you are very much. I anticipated that, as I anticipated the coming of a religion founder — such is a certain behaviour of movements which are based on empiricism. The analogy with Comte's positivism is not so far away. But why not — we are mortals, and therefore we have to be like mortals. [WKA (1943a)][24]

Carnap attempted to counter the criticism that Neurath directed at his position, regretting that "you are so intolerant with respect to some ideas of people who share with us the fundamental empiricist attitude." [WKA (1943b)] In fact, in his reply Carnap was effectively bringing forth his notion of tolerance (for the first time in connection with the semantic issues), and making remarks about the general nature of philosophical concept

[23]Cited with the compliance of the *Wiener Kreis Stichting* (Amsterdam). All rights reserved.

[24]Cited with the compliance of the *Wiener Kreis Stichting* (Amsterdam). All rights reserved.

formation. These remarks are already imbued with the fundamental *ethos* of the idea of explication:

> In the case of Popper, I believe your reaction is chiefly caused by the fact that he criticized the Vienna Circle quite unnecessarily. He was overcritical and so you are now. Even when he wrote the book [Logik der Forschung] he was in agreement with us on most fundamental points. When later he came into personal contact with us the agreement became even more strong and conscious to him. Some of his views which you criticize, e.g. the refutability of hypotheses, have the same defects — and, I think, the same merits — as many of our earlier views: *they might be taken as first approximations but closer inspection shows that they are not entirely adequate but must be replaced by better approximations.* I suppose that the same holds for many of our present views, *including mine*, where we do not see today how they should and can be improved. [...] I think for the sake of the movement it would be much better if we were more tolerant towards each other. If your intolerance would become the general custom, then I am afraid that you would be among the first to be declared a heretic and excommunicated. By tolerance, of course, I do not mean acceptance of each others' views. The differences of opinion should and will be discussed. But this discussion is not helped by labeling the views of the others as nonempiricist and metaphysical. [WKA (1943b)][25]

Carnap is here quite serious about the necessity of propounding more generally the conversational maxims pertaining to the Principle of Tolerance. However, the full content and the implications of the principle are discussed but in a cursory manner. It would surely have benefited Carnap in his juxtaposition with Neurath (and more generally in the realization of his overall program of explication) if he had more explicitly adduced the presuppositions of the notion of tolerance that he leaned on. What is especially damaging to Carnap's program of explication is that the *moral* presuppositions of tolerance (besides purely *logico-mathematical* aspects) were never made fully explicit by him which also renders it understandable why Neurath had difficulties in grasping his position and why the overall significance of the Principle escaped so many of his colleagues. At the end of his letter Carnap returns to the more technical discussion of the *Semantics* and reminds Neurath that "when you read my 'Semantics' and especially when you write to me about it, please keep in mind that the semantical concepts used there are meant for application also to the language of science, especially the concepts 'true', 'L-true', and similar ones." [*ibid.*] Indeed, Carnap is here reminding Neurath of the ultimate motivation of his own work that is still, despite all the skepticism of Neurath, thoroughly based on the wish to construct concepts that could be of concrete use to scientists in their cognitive enterprise. It seems that Neurath here completely misses the more nuanced and relaxed conception of language-engineering that Carnap has already envisaged and is effectively implementing: "In any case, the concept of truth as I deal with it is meant as a systematization of the inexact term 'true' as used by scientists and in everyday life." [*ibid.*]

[25]Cited with the compliance of the *Wiener Kreis Stichting* (Amsterdam). All rights reserved.

5.2.2 A summary of Carnap's work in semantics in the 1940s

"I do not indulge in this vice [modal logic] generally and thoroughly. [...] Although we do not like to apply intensional languages, nevertheless I think we cannot help analyzing them. What would you think of a entomologist who refuses to investigate fleas and lice because he dislikes them?" — Carnap to Quine in 1938

Carnap first presented the new semantical viewpoint in *Foundations of Logic and Mathematics* [Carnap (1939)], where a "logical calculus" is distinguished from other calculi simply in virtue of its "customary interpretations". Carnap writes that "[t]his classification is rather rough and is only meant to serve a temporary, practical purpose." [*ibid.*, 29] However, Carnap does not provide an improved or less "rough" account of this classification later in the monograph. Within just a few years Carnap embarked on a larger project of explicating semantic notions. The result was *Studies in Semantics*, the first volume of which was *Introduction to Semantics* (1942), and the second *Formalization of Logic* (1943).[26] Finally, an account of semantics built on a new basis, *Meaning and Necessity*, was published in 1947. In this book, Carnap made heavy use of the distinction between extension and intension, utilizing modal notions. Carnap gave a succinct description of the emergence and content of the semantic viewpoint in his philosophy in the preface to the *Introduction to Semantics*:

> Tarski, both through his book and in conversation, first called my attention to the fact that the formal method of syntax must be supplemented by semantical concepts, showing at the same time that these concepts can be defined by means not less exact than those of syntax. Thus the present book owes very much to Tarski, more indeed than to any other single influence. On the other hand, our conceptions of semantics seem to diverge at certain points. First [...] I emphasize the distinction between semantics and syntax, i.e., between semantical systems as interpreted language systems and purely formal, uninterpreted calculi, while for Tarski there seems to be no sharp demarcation. Second, within semantics, I stress the distinction between factual truth, dependent upon the contingency of facts, and logical truth, independent of facts and dependent merely on meaning as determined by semantical rules. I believe that this distinction is indispensable for the logical analysis of science; and one of the chief problems discussed in this book is that of representing this distinction, which has been made in some form or other by most philosophers since ancient times, by exact semantical definitions. Here again, Tarski seems to doubt whether there is an objective difference or whether the choice of a boundary line is not more or less arbitrary. (The two points of divergence mentioned seem, incidentally, to go back to a common root, namely to the distinction between logical and descriptive signs.) At present, it is not quite clear to me whether the divergence is a genuine difference of opinion ir per-

[26] Incidentally, the second volume was written before the first one.

haps merely a difference in emphasis, direction of attention, and preference in procedure. [Carnap (1942), x–xi]

As it turned out, Carnap actually failed to answer satisfactorily the question about the objectivity of the distinction: "Our previous discussion has shown the difficulties connected with the problem of a general formulation of these distinctions (§§13 and 16). This problem is very much in need of further investigation.". [Carnap (1942), 243] More generally, as Steve Awodey has pointed out, there are several serious problems afflicting Carnap's semantic program for defining analyticity (notwithstanding the earlier liberating result found during the correspondence with Gödel): "First, despite his efforts in *Formalization of Logic*, Mac Lane's problem of capturing the distinction between logical and descriptive symbols in a general way still remained:"[27] [Awodey (2007) [Friedman & Creath (2007), 236]]

> So far we have discussed the distinction between logical and descriptive expressions only in the form in which it appears when we have to do with a particular semantical system, in other words, as a question of special semantics. The problem is more difficult in the form it takes in *general semantics*. Here it is the question whether and how "logical" and "descriptive" can be defined on the basis of other semantical terms, e.g., "designation" and "true", so that the application of the general definition to any particular system will lead to a result which is in accordance with the intended distinction. A satisfactory solution is not yet known. [Carnap (1942), 59]

"Second," Awodey continues, "how do we explain the all-important idea that the truth of a sentence 'follows from the semantic rules alone'?" [*ibid.*] Whereas in the *Logical Syntax* Carnap's idea had been to identify the logical truths with the semantic truths consisting entirely of logical symbols, the same strategy could not be implemented in the framework of the *Studies in Semantics*. Because in that context *all* terms are interpreted, the former procedure became too narrow. Consider the example provided by Awodey (where S denotes the predicate "is blue" and a is an individual constant denoting Titisee[28] [*ibid.*]:

$S(a)$	true, because the Titisee is blue;
$\neg(S(a)\&\neg S(a))$	true, but not because of anything about the Titisee

Carnap's idea here is that the truth of the second clause holds independently of of the

[27] "Mac Lane's problem" refers to the objection, raised by Saunders Mac Lane in his critique of Carnap's *Syntax*, that Carnap's definition of the logical symbols as the largest collection of symbols such that every sentence constructed only from them is determinate on the basis of transformation rules, is mathematically unworkable. The gist of the objection is that there is no *unique* such maximal set of symbols. For, as Awodey makes clear, "while there are various different sets that are maximal, in the sense that they admit no further extension, the intersection of all such sets (as Carnap proposes) is itself no longer maximal. The situation is similar to one arising frequently in abstract algebra, as Mac Lane surely recognized." [Awodey (2007), 233n] Mac Lane's criticism was published as [Mac Lane (1938)].

[28] I.e., a lake in the southern Black Forest in Baden-Württemberg.

interpretation of the *non-logical* constants S and a occurring in it, and thus independently of the (physical) fact that the Titisee is blue. What is characteristic of the second clause that it remains true, even if we substitute entirely different non-logical constants for S and a, e.g., b for a, or $R(x, c)$ (or even $Q(x_1, \ldots, x_n)$) for S and so on. The putative definition of a sentence that is true by the sole meaning of the words that it contains *could* then be stated as: the analytic sentences are precisely the ones that remain true for all substitutions for their *non-logical* constants. This does not amount to an improvement, however, because it presupposes the distinction between logical and descriptive symbols which was precisely seen to be the crux of the "Mac Lane problem" referred to above. Another important criterion for analytic sentences to be characterized in the semantic framework is that they should be devoid of content. Thus, it becomes obligatory to demonstrate that logical truth does not depend on or imply factual (contingent) truth. As Awodey puts it, in the *logical Syntax* this was to have been ensured by "(1) the prohibition of extra-logical 'meanings' combined with (2) the requirement of determinacy: analyticity is then empirically empty, since (by (1)) no empirical facts are used in its specification, and (by (2)) any consequence of an analytic proposition is analytic." [*ibid.*, 237] What complicates the situation in the context of *Studies* is the fact that Carnap works with a mixed framework of both logical and empirical concepts and propositions. Combined with the use of "extra-logical" meanings such a system is especially conducive to the risk that the two might become entangled. A purportedly promising line of attacking this problem would be to stipulate that every purely mathematical sentence is "semantically determinate" in the sense that it would be uniformly true (false) under all *permissible* substitutions. In that case its content (defined as its consequence class) would be trivial.

Here one encounters the fundamental shortcoming of Carnap's account of semantics. The problem for Carnap lies mainly in the fact that if one is to determine logical truth (nowadays called logical validity), the substitutional strategy of varying the valuations of the different constant symbols in a *single* interpretation is not, generally, sufficient. What Carnap would essentially need is the modern model-theoretical characterization of logical truth: the idea that the truth of a logically true (valid) sentence is independent of the interpretations of its non-logical constants in *all possible* interpretations of these constants over *all possible* domains of quantification. Thus, he lacks the distinction between the notions of *truth in a particular model* and *truth in all models*. However, the model-theoretical viewpoint was simply absent in the 1940s. Even Tarski, who is generally considered to be the pioneer in this field, did not have recourse to purely model-theoretical concepts in his classic paper on truth [Tarski (1936)], nor in his other works until 1952.[29] In keeping with this, Carnap's semantical work in the 1940s always took place within a framework of a language with a single fixed interpretation. This is a general feature of Carnap's work in semantics, an *idée fixe*, which Hintikka has called the *one-domain assumption* (shared by Carnap and Wittgenstein):

> This one-domain assumption is probably the most important background assumption in Carnap's later philosophy of logic and, to some extent, also in his

[29] As Awodey remarks, "the idea seems first to have suggested by Kemeny [Kemeny (1948)], who was an associate of Carnap and was explicitly responding to Carnap's semantic work." [*ibid.*, 238n]

> philosophy of mathematics. Its manifestations and consequences are worth spelling out somewhat more fully, partly for the purpose of illustrating the import of this assumption.
>
> First, we can register a clear-cut consequence of the one-domain assumption in Carnap's approach to the foundations of mathematics. If one assumes, as Carnap does, that each interpreted (first-order) language carries with itself a fixed given domain of individuals, one cannot compare in one's language tow different domain with each other. Yet, such comparisons played an interesting role in the background work in the so-called extremality axioms, that is axioms calculated to enforce the maximality or minimality of the domain of the intended models of a mathematical axioms system. [...]
>
> The same assumption was ingrained elsewhere in Carnap's thinking. It clearly is a relic of the universalist attitude, and in a certain sense it rules out, or at least discourages, a consistently model-theoretical attitude. For Carnap, the semantical value of an individual constant is an individual. The main problem that an applier of semantics faces is then to keep track of that given individual as it enters and exits the different possible configurations of such given individuals. For a consistent model theorist, the prime given materials in one's semantics are the possible worlds themselves. An individual is conceived simply as the function that picks out as its value that particular individual in the different worlds in which it can make its appearance. And this function is, in principle, chosen by us — not, of course by each of us individually but by the tacit decisions of the language community. These decisions are codified in what has been called [...] individuating functions. [Hintikka (1997), 200–201]

This evaluation of Carnap's conception of semantics is poignant. As Hintikka has elsewhere emphasized [Hintikka (1975a)], Carnap was very much in possession of the basic concepts needed to take the steps towards the fully-fledged model-theoretical viewpoint, but that against all odds, he did not take these steps. This is all the more curious as Carnap was in possession of these tools (or similar ones) already in 1927–1930 when he was occupied writing his *Untersuchungen zur Allgemeinen Axiomatik*. In that work he pioneered the notion of a "model" of a formal language, which he used to define the notion of logical entailment — "A implies B" — along the lines of "every model of A is a model of B".[30] Psychologically, the cautiousness about employing these tools later might have been affected by the overall failure of that work the goal of which, as we remember, was to establish a general framework for metalogic, including the specification of the relationship between the different notions of completeness. Finally, it was precisely Carnap's own conception of entailment that logicians recognized as the required notion to build the model-theoretical notion of logical consequence on and which they were able to recognize as providing the solution to the problems that Carnap was tackling with in his *Studies in Semantics* and in the later *Meaning and Necessity*.

[30] "Wir hatten g eine 'Folgerung' von f genannt, wenn $f \to g$, d.h. $(\mathcal{R})(f\mathcal{R} \to g\mathcal{R})$ gilt; diese 'Implikationaussage von g' kann jetzt so ausgesprochen werden: '**alle Modelle von** f **sind auch Modelle von** g'." [Carnap (2000), 95]

Following his work in the *Studies in Semantics* Carnap gradually worked his way towards implementing the strategy of distinguishing between analytic and synthetic sentences by having recourse to more precisely specified version of intensional identity and logical "necessity" (with empirical "contingency" corresponding to syntheticity). First in the paper "Modalities and Quantification" [Carnap (1946)], and then more explicitly in *Meaning and Necessity* [Carnap (1947)], he developed a series of intensional and modal languages which constituted a framework for proposing a new definition of analyticity, i.e., analyticity defined in terms of intensionality or logical necessity (instead of truth). According to Carnap's new proposal, the notions of intension and logical truth can be related as follows [Awodey (2007), 241]:

Two expressions A and B *have the same extension* iff $A = B$ is true;
they are then said to be *extensionally equivalent*.

Two expressions A and B *have the same intension* iff $A = B$ is L-true;
they are then said to be (*intensionally*, or) *L-equivalent*.

This distinction is further refined by the following definitions provided by Carnap.[31] First, the one given for *extensional contexts*: [Carnap (1947), 48]

Definition 5 a. *The* expression $\mathfrak{A}_i$ *is* ***extensional*** with respect to a certain occurrence *of* $\mathfrak{A}_j$ *within* $\mathfrak{A}_i$ *in the system* $S =_{Df}$ $\mathfrak{A}_i$ *and* $\mathfrak{A}_j$ *are designators; the occurrence in question of* $\mathfrak{A}_j$ *within* $\mathfrak{A}_i$ *is interchangeable with any expression equivalent to* $\mathfrak{A}_j$ *in S.*

b. *The* expression $\mathfrak{A}_i$ *is* ***extensional*** *(in S)* $=_{Df}$ $\mathfrak{A}_i$ *is a designator (in S);* $\mathfrak{A}_i$ *is extensional with respect to any occurrence of a designator within* $\mathfrak{A}_i$ *(in S).*

c. *The* semantical system *S is* ***extensional*** $=_{Df}$ *every sentence in S is extensional.*

Second, the one for *intensional contexts*: [*ibid.*]

[31] The distinction between intension and extension that Carnap introduced in *Meaning and Necessity* was not at all new. It had a venerable history with its roots traceable back at least to Aristotle, although the literal use of terms having the same function derives from the early modern period. Perhaps the most natural point of reference is the *Logique de Port Royal* of Pierre Nicole and Antoine Arnauld, the two sharp-witted Jansenists. The following table gives a rough picture of the development of the terminology and the use of the distinction among some of the most prominent philosophers of the last 350 years. (The table is adapted from [Krauth 1970, 37].)

Logique de Port Royal:	compréhension – extension
J. S. Mill:	connotation – denotation
Frege:	Sinn – Bedeutung
Russell:	meaning – denotation
Quine:	meaning – reference
Church:	sense – denotation
Carnap:	intension – extension (nominatum)

Definition 6 a. *The* expression $\mathfrak{U}_i$ *is* ***intensional*** with respect to a certain occurrence *of* $\mathfrak{U}_j$ *within* $\mathfrak{U}_i$ *in the system* $S =_{Df}$ $\mathfrak{U}_i$ *and* $\mathfrak{U}_j$ *are designators;* $\mathfrak{U}_i$ *is not extensional with respect to the occurrence in question of* $\mathfrak{U}_j$ *within* $\mathfrak{U}_i$*; this occurrence of* $\mathfrak{U}_j$ *within* $\mathfrak{U}_i$ *is L-interchangeable with any expression L-equivalent to* $\mathfrak{U}_j$ *in S.*

b. *The* expression $\mathfrak{U}_i$ *is* ***intensional*** *(in S)* $=_{Df}$ $\mathfrak{U}_i$ *is a designator;* $\mathfrak{U}_i$ *is, with respect to any occurrence of a designator within* $\mathfrak{U}_i$*, either extensional or intensional, and is intensional with respect to at least one occurrence of a designator.*

c. *The* semantical system *S is* ***intensional*** $=_{Df}$ *every sentence in S is either extensional or intensional, and at least one is intensional.*

Simple examples of this distinction could be devised as follows: let us consider a propositional function having the form "x is y", where x ranges over names or definite descriptions, and y ranges over properties. Then, the context "x is white" is extensional: e.g. "the sixth tallest building in Helsinki is white" follows on the basis of extensional equivalence from "The Tower of the Olympic Stadium in Helsinki is white". By contrast, "I believe x is white" is intensional , because "I believe the Tower of the Olympic Stadium in Helsinki is white" follows from "I believe white is the colour of the Tower of the Olympic Stadium in Helsinki", but not from "I believe the sixth tallest building in Helsinki is white". The distinction is applied by Carnap to solve puzzles like the "antinomy of identity", which is epitomized by Frege's famous example of Venus. Clearly, it is necessarily true that "Phosphorus is Phosphorus", but it requires empirical inquiry to find out that "Phosphorus is Hesperus". In the latter sentence, the extensions of the names are the same, viz., the planet Venus, but the intensions differ, the one signifying the notion "morning star" and the other the notion "evening star". A considerable systematic advantage of Carnap's system in the *Meaning and Necessity* is that not only are the notions of intension and extension definable from L-truth, but also they and the alethic modalities are all interdefinable. For example,

$N(A)$ is true iff A is intensionally equivalent to *True*
(which holds iff A is L-true)

from which it follows:

$N(A = B)$ is true iff A is intensionally equivalent to B
(i.e. $A = B$ is L-true).

The philosophical import of these definitions is that the three distinctions that play such a prominent role in Carnap's theory of semantics, viz., intension vs. extension, analytic vs. synthetic, and necessity vs. possibility "become just three sides of the same coin, as it were." [Awodey (2007), 242] We have already alluded to the characteristic difficulties inherent in the Carnapian approach. Lacking a genuinely model-theoretical point of view Carnap did not grasp the practical possibility of devising an extensional semantics for modal languages. This realization had to wait for the logical results independently found by Jaakko Hintikka (1957), Stig Kanger (1957) and Saul Kripke (1959).[32] By this

[32]Steve Awodey mentions in this context only Saul Kripke's work of 1963 [Kripke (1963)], betraying an

time, however, Carnap's interests already lay elsewhere. From 1945 on he is increasingly occupied with the work on inductive logic and foundations of probability. This stage would then last until his death in 1970. Summing up his work in semantics, Hintikka writes:

> [I]t seems to me that possible-worlds semantics overwhelmingly suggests that Carnap was on the right track. It makes the weight of his reply to Quine (that is what "Meaning and Synonymy in Natural Languages" [Carnap (1955a)] essentially is) felt in a new way, and puts the onus of producing specific criticisms much more on Carnap's critics than has been recognized in recent discussion. What is even more important, it suggests new constructive, empirical approaches to the pragmatics of beliefs and intensions. As such, it amounts to an important partial vindication of Carnap vis-à-vis his critics, and shows the power of his ideas to inspire and to guide further development of the studies to which he himself already contributed so much. [Hintikka (1975a), 98]

insufficient grasp of the central developments in the field before the 1960s.

Chapter 6

The Ideal of Explication as an Embodiment of Rationality

6.1 Explication in its Context: *The Logical Foundations of Probability*

During the years 1942–1944, having been granted a leave of absence by the University of Chicago and financed by the Rockefeller Foundation, Carnap developed many of the features of the theory of inductive logic (including results concerning the central conceptual tool of that theory, viz., the function $\mathfrak{c}^*$). These investigations finally resulted in a formulation of a system of inductive logic based on a *logical* notion of probability, an account of which was ultimately published as the monograph *The Logical Foundations of Probability* in 1950. [Carnap (1950b)] This monograph was planned only as the first volume in a projected two-volume work called *Probability and Induction*. In this first volume Carnap set out to tackle the fundamental problem of formulating a new basis for a theory of probability which would account for the various uses of the notion "probability" in science as well as in everyday-life. According to Carnap, "[o]ne of the main tasks of any new theory of probability is to supply adequate explicata for the concept of probability and the for the methods of inductive reasoning which are at present applied in science and in statistics. However, there does not seem to be sufficient clarity and agreement concerning the requirements that an adequate explicatum for any explicandum must fulfill." [Carnap (1950b), vii] To compensate for this state of affairs, Carnap provides in the first chapter of the book a methodological discussion about explication, "although", he goes on to remark, "this topic should be dealt with more appropriately in a book on concept formation in science." [*ibid.*] In spite of these reservations, the first chapter of the book is worth of serious consideration; it remained the most elaborate discussion about the method of explication Carnap ever provided. In its scant 18 pages it intends to shed some light on the radically new framework of concept formation that Carnap had already envisaged in the 1940s in connection with his work on semantics.[1] The opening passage of §2 provides the first tentative description of the new method:

> The task of **explication** consists in transforming a given more or less inexact concept into an exact one or, rather, in replacing the first by the second. We call the given concept (or the term used for it) the **explicandum**, and the exact concept proposed to take the place of the first (or the term proposed for it) the **explicatum**. The explicandum may belong to everyday language or to a previous stage in the development of scientific language. The explicatum must be given by explicit rules for its use, for example, by a definition which incorporates it into a well-constructed system of scientific either logicomathematical or empirical concepts. [*ibid.*, 3]

Carnap then goes on to relate the historical background of the notion in the works of Kant, Husserl and C.H. Langford. Kant's notion of explication refers here to the process of obtaining a particular predicate (a concept) by analysis of the subject (an individual concept). This is inextricably entwined with the Kantian notion of analyticity which we presented in the Chapter 3. Kant seems to understand explication in the sense of clarifying what is already *implicitly* thought in the subject, thereby making explicit the *conceptual content* of an initially non-determinate (individual) concept through an explicative judgement [*Erläuterungsurteil*]. In the same paragraph Carnap implies that his notion of explication is partly motivated by Husserl's talk of "explication" as "the synthesis of identification between a confused, non-articulated sense and a subsequently intended distinct, articulated sense" the latter of which Husserl calls the "Explikat" of the former. [*ibid.*] Although Carnap provides in this connection as his only reference the [definition by Dorion Cairns in the] *Dictionary of Philosophy* [1942] edited by D. Runes, it is doubtful that this is *the* source of Carnap's knowledge about Husserl's conception.[2] Although the most conspicuous signs of Husserl's influence on Carnap can be detected in Carnap's early works, especially in *Der Raum* and in the *Aufbau*, it would seem rather odd if Carnap did not keep himself informed about the evolution of Husserl's views. Although no direct influence can here be detected, it still serves a definite purpose to adduce Husserl's notion of explication as a backdrop for the following presentation. Husserl's most substantial discussion on explication occurred in *Experience and Judgement* [*EJ*], in chapter 2 of Part I. This work was edited by Ludwig Landgrebe (Husserl's assistant at the Freiburg University) under the authorization of Husserl, and published posthumously in 1939 (Husserl died in 1938). The central concepts that figure in Husserl's distinction are, on the one hand, the notion of "simple" or "immediate" apprehension [*schlichte Erfassung*], and on the other hand, the notion of "Explication" [*Explikation*], which can enter into the

[1] Very early in *Meaning and Necessity* (§2) Carnap writes: "The task of making more exact a vague or not quite exact concept used in everyday life or in an earlier stage of scientific or logical development, or rather of replacing it by a newly constructed, more exact concept, belongs among the most important tasks of logical analysis and logical construction. We call this the task of explicating, or of giving an *explication* for, the earlier concept; this earlier concept, or sometimes the term used for it, is called the *explicandum*; and the new concept, or its term, is called an *explicatum* of the old one." [Carnap (1947), 8–9]

[2] Michael Beaney provides no evidence to back up his belief that this indeed is the case. [Beaney (2004), 141]

"synthesis of identification". [*EJ*, §22] The distinction sets a coördinate frame for the phenomenological analysis of the different levels of reflective perception of an object. These levels constitute a hierarchy in which the "simple apprehension" is the lowest level at which we are conscious of the object as a whole deprived of any particular characteristics. Transgressing the boundaries that delimit the core of such a qualitatively simple content of consciousness is possible through the re-cognition of the "internal horizon" associated with every experience of a thing. Husserl had asserted the existence of such a "horizon" in [*EJ*, §8], where its significance was explained to be that of defining the boundaries of an area of possible knowledge regarding a particular object. Two other important concepts that characterize the qualitative content of our experience, according to Husserl, are the "retentional" and "protentional" aspects of the structure of experience. Strikingly in line with the modern view of cognitive psychology, Husserl explains that these aspects are exemplified in the way that our experience is informed ('modulated') by our existing knowledge and expectations. We may recall previous perceptions (as in an auditory retention of a chord quality) or already have a clear comprehension of the *type* that a particular object of our perception instantiates ("*That* is a dominant-seventh altered chord with a 13th."). Concerning visual perception, for example, we can *imagine* what the object would look like from a different point of view in space or anticipate how its color changes if we move to another position (taking into account the reflection of light from a stained surface, for example). The level of *explication* in the Husserlian picture of perception marks the enrichment of our knowledge as we elucidate more and more detailed aspects of the object:

> *Explication is penetration of the internal horizon of the object by the direction of perceptual interest.* In the case of the unobstructed realization of this interest, the protentional expectations fulfill themselves in the same way; the object reveals itself in its properties as that which it was anticipated to be, except that what was anticipated now attains original givenness. A more precise determination results, eventually perhaps partial corrections, or — in the case of obstruction — disappointment of the expectations, and partial modalization. [*EJ*, §22]

What should especially be noticed is that Husserl here distinguishes explication from "analytic clarification" [*analytische Verdeutlichung*], where the object of the explication need not be "intuitively given". An example is the analysis of "brother" as "male sibling" (Chico does not have to think about Harpo, Groucho, Gummo or Zeppo to perform the analysis). Husserl clearly does not have conceptual analysis (of the sort of Kant, for example) in mind as he presents the distinction. Rather, the Husserlian explication is an enrichment of sense that operates *within* the domain of intuition. But as the above passage makes clear, Husserl allows for partial corrections in the explication, if our expectations concerning perception are disappointed. This aspect comes quite close to the Carnapian idea of explication. In this sense, *both* Carnap and Husserl see the explication as a process or vehicle of *precisification* by means of which, as Beaney puts it "our ordinary understanding is refined, and if necessary, transformed." [Beaney (2004), 142] Carnap, however, sees his own view of explication as a relative of C. H. Langford's conception of analysis: "What I mean by 'explicandum' and 'explicatum' is to some extent similar to what C. H. Langford calls 'analysandum' and 'analysans': 'the analysis then states an appropriate relation of equivalence between the analysandum and the analysans'. [...]; he says that the motive of an analysis 'is usually that of supplanting a relatively vague idea by a more precise one'." [*ibid.*, 3] However, Carnap remarks that his procedure of explication must be understood in a wider sense than the (traditional) procedures of analysis or clarification of Kant, Husserl and Langford. Indeed, what he wants to emphasize is that his notion of explication does not fully comply with the traditional notion of philosophical analysis (in its decompositional and resolutive senses), but comprises rather a different (transformative) kind of meaning conferring: "The explicatum (in my sense) is in many cases the result of an analysis of the explicandum (and this has motivated my choice of the terms); in other cases, however, it deviates deliberately from the explicandum but still takes its place in some way [...]." [*ibid.*]

The above account of the historical and motivational background of Carnap's method of explication captures only some of the aspects that are relevant for characterizing its overall significance. In the following sections we will see how the method can be understood in a wider frame, providing an account of what may be termed the "ideal of explication". The thesis that Carnap's method, conceived as a regulative ideal of assessing different proposals for theoretical frameworks or languages to be used in different domains of human activity, is to be taken as a formulation of a modern conception of rationality or reasonableness, is more closely scrutinized. I will especially concentrate on the suggestion, systematically developed by A. W. Carus [Carus (2007)], that Carnap's method of explication embodies a distinctively modern variety of Enlightenment thought. Now, before I proceed to present the wider implications of Carnap's method and assess the fortunes of the interpretation of Carus, it is necessary to provide a more accurate account of the *desiderata* that Carnap *himself* set for the explicata to be worked out in the different situations. This will be the subject of the next section.

6.1.1 The criteria for explication

To begin with, it is necessary to clarify the meaning of the term "explication" on a general level so as to exclude the possibility of any misunderstanding. To provide a clear-cut starting point for our discussion, consider the definition of "explication" provided in the *Chambers' Etymological Dictionary*. According to this source, the use of *explication* as an English term is documented from 1528 on as having the meaning "detailed statement or account [...] from *explicāre* unfold". [3] Regarded as a special philosophical term "explication" does not refer *solely* to the elimination of conceptual ambiguities or vagueness, contrary to what is often thought to be the case. Were explication understood merely in this sense, it would be tantamount to the method of analysis to which it admittedly bears a close relationship. Explication, as Carnap conceives it, comprises essentially also a *synthetic* ingredient, as we already noticed above. According to Carnap's crucial insight,

[3][*Chambers Dictionary of Etymology*, Chambers Harrap Publishers (1988)]

explication must be regarded as an instrument for comparing and evaluating *entire artificial language systems*. The practical task of comparison and evaluation is naturally most often undertaken in a stepwise manner, one concept at a time, especially when one of the systems under consideration is our colloquial language. But in the case of artificial and highly technical language systems (such as the various systems of intensional logic) we have at our disposal powerful meta-theoretical tools to implement a much more efficient comparison and evaluation. As far as these more technical contexts of explication are concerned, I will present some examples relating to mathematical physics in the sections to come. For now it suffices to clarify the basic features of explication as Carnap presents them in the *LFP*.

The conspicuous feature of Carnap's account of explication is the asymmetry between the *explicandum* and the *explicatum*. While the former is a non-articulate and inexact, sometimes even vague, concept (possibly taken from everyday-language), the latter is articulate, exact and distinct. As Carnap readily remarks, one must define certain conditions on how the problem itself ought to be stated, i.e., how the explicandum should be given:

> There is a temptation to think that, since the explicandum cannot be given in exact terms anyway, it does not matter much how we formulate the problem. But this would be quite wrong. On the contrary, since even in the best case we cannot reach full exactness, we must, in order to prevent the discussion of the problem from becoming entirely futile, do all we can to make at least practically clear what is meant as the explicandum. What X means by a certain term in contexts of a certain kind is at least practically clear to Y if Y is able to predict correctly X's interpretation for the most simple, ordinary cases of the use of the term in those contexts. It seems to me that, in raising problems of analysis or explication, philosophers very frequently violate this requirement. They ask questions like: 'What is causality?', 'What is life?', 'What is mind?', 'What is justice?', etc. Then they immediately start to look for an answer without first examining the tacit assumption that the terms of the question are at least practically clear enough to serve as a basis for an investigation, for an analysis or explication. [Carnap (1950b), 4]

The task of explication must, then, be preceded by the task of *clarification*. Those involved in the process must reach a satisfactory degree of agreement about what they are *de facto* explicating. Carnap is optimistic about the prospects of achieving the requisite mutual understanding: "Even though the terms in question are unsystematic, inexact terms, there are means for reaching a relatively good mutual understanding as to their intended meaning. An indication of the meaning with the help of some examples for its intended use and other examples for uses not now intended can help the understanding. An informal explanation in general terms may be added. [...] By explanations of this kind the reader may obtain step by step a clearer picture of what is intended to be included and what is intended to be excluded; thus he may reach an understanding of the meaning intended which is far from perfect theoretically, but may be sufficient for the practical purposes of a discussion of possible explanations." [*ibid.*, 4–5] In spite of this optimism, Carnap very clearly perceived the challenge of the largely *informal* task of establishing the identity of the explicandum that should precede the process of replacement with a more exact concept. Another, perhaps more serious, problem arises in the situation when the explicandum, informally identified, is chosen from an evolved, 'natural' language, and juxtaposed with an explicatum precisely defined within a language having more artificial or constructed features. The relation between such concepts cannot be precise. This has generally been taken as a reason for not taking Carnap's suggestions seriously at all. Regarding the method of explication as a form of *analysis*, the critics take the non-specificity of the relation between the explicandum and explicatum as a decisive reason for rejecting the method out of hand. What these critics very much ignore is that the framework of explication is diametrically opposite to the earlier program of rational reconstruction in terms of its domain of application. Whereas the method of translation into the 'formal mode of speech' had been the characteristic of the pre-1931 rational reconstruction (i.e. before the *Syntax*), the process of translation being internal to the language, the method of explication was presented explicitly as a process operating externally to the precise target language in which a particular explication was to be adduced. As Carnap remarks:

> Since the datum is inexact, the problem itself is not stated in exact terms; and yet we are asked to give an exact solution. This is one of the puzzling peculiarities of explication. It follows that, if a solution for a problem of explication is proposed, we cannot decide in an exact way whether it is right or wrong. Strictly speaking, the question whether the solution is right or wrong makes no good sense because there is no clear-cut answer. The question should rather be whether the proposed solution is satisfactory, whether it is more satisfactory than another one, and the like. [Carnap (1950b), 4]

The predicator "right or wrong", of course, means "true or false" in Carnap's locution; and the predicator "satisfactory" is understood in the sense of "being useful or fruitful for the practical purposes in question". This fundamentally *pragmatic* ingredient of the method of explication as well as the consequent distinction between the *internal* and *external* questions[4] pertaining to a language or a group of languages was clearly pointed out by Carnap's student Howard Stein:

> The explicatum, as an exactly characterized concept, belongs to some formal-

[4]This distinction is figured also in a completely different vein, i.e. in Pierre Bordieu's conception of science as 'fields of activity'. According to Bordieu, the goal of the scientist is to achieve recognition from competitor peers. This primary objective, however, does not preclude the possibility of the scientist furthering the progress of science itself, as is made clear in [Bordieu (1975)]. The progress of science is dictated by the dialectical relation between the internal and external determinants which are constituted, respectively, by the scientific field and by what Bordieu calls "habitus". [Lloyd (1993)] gives a succinct definition of this notion: "The habitus is the principle of a form of subjective, implicit, practical knowledge that is unconscious and not requiring consciousness, although it masters objective necessity, thus enabling social competence." [*ibid.*, 33] In Bordieu's account, as Lloyd relates, the subjective/objective dichotomy in explanation is overcome through the realization of the 'ontological complexity' of the two determinants.

> ized discourse — some 'framework'. The explicandum [...] belongs ipso facto to a mode of discourse outside that framework. Therefore *any* question about the relation of the explicatum to the explicandum is an 'external' question; this holds, in particular, of the question whether an explication is adequate — that is, whether the explicatum does in some appropriate sense fully represent, within the framework, the function performed (let us say) 'presystematically' by the explicandum. [Stein (1992), 280]

One can here discern the characteristically Carnapian conception of a mutual feedback relation between the different linguistic frameworks, one of which functions as the medium for identifying an *explicandum* through a process of clarification, and the second of which is the 'more constructed' medium for providing an *explication* of the *explicandum*. The external question concerns, essentially, "the practical problem whether or not to incorporate into the language the new linguistic forms" pertaining to the practical situation at hand. [Carnap (1950a), 209] Now, provided that the participants of a discussion concerning the adoption of a suitable linguistic framework for a specific use have a preliminary mutual understanding of the concepts to be specified (via the precisification enabled by explication), what are the general conditions that must be imposed upon the explicata themselves? Carnap answers this question in §3 of *LFP* titled "Requirements for an Explicatum": "If a concept is given as an explicandum, the task consists in finding another concept as its explicatum which fulfills the following requirements to a sufficient degree.": [(1950b), 7]

1. The explicatum is to be *similar to the explicandum* in such a way that, in most cases in which the explicandum has so far been used, the explicatum can be used; however, close similarity is not required, and considerable differences are permitted.

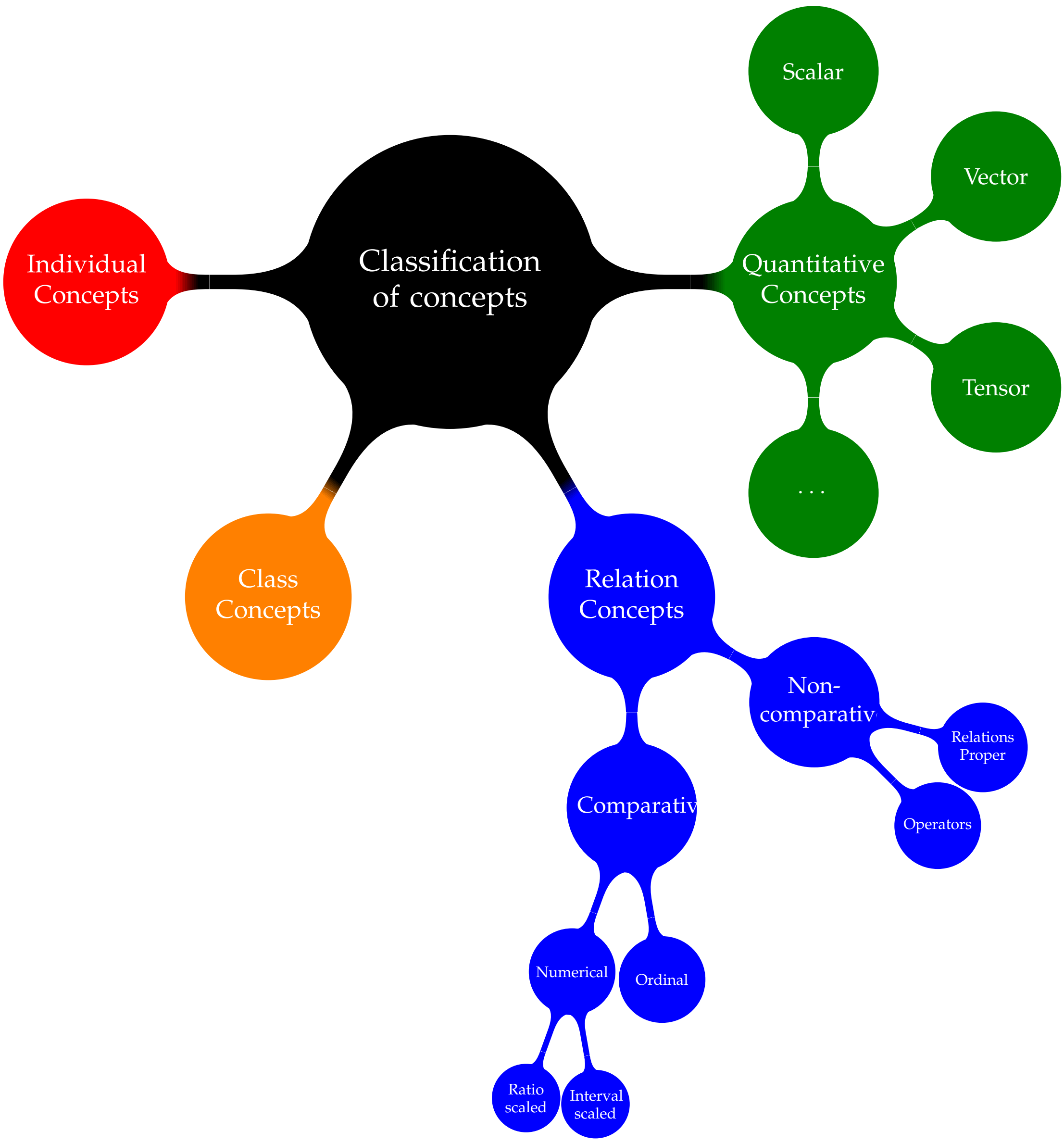
2. The characterization of the explicatum, that is, the rules of its use (for instance, in the form of a definition), is to be given in an *exact* form, so as to introduce the explicatum into a well-connected system of scientific concepts.

3. The explicatum is to be a *fruitful* concept, that is, useful for the formulation of many universal statements (empirical laws in the case of a nonlogical concept, logical theorems in the case of a logical concept).

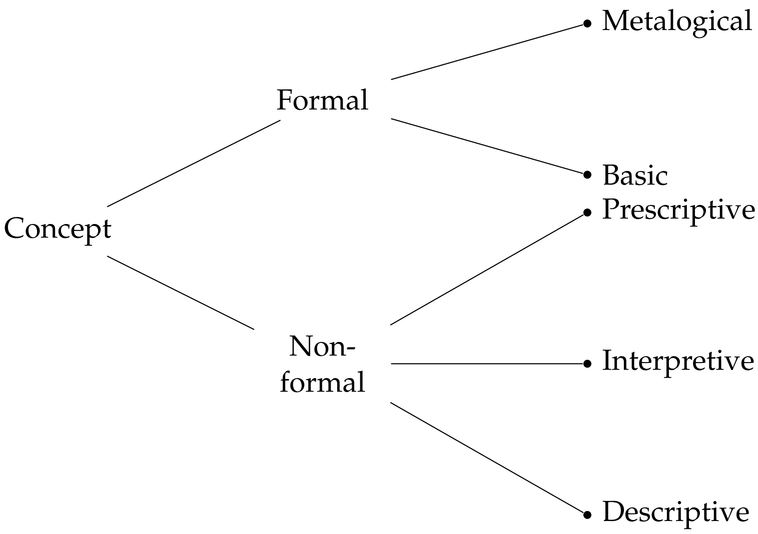
4. The explicatum should be as *simple* as possible; this means as simple as the more important requirements (1), (2), and (3) permit.

As Carnap makes clear in (4), the requirements are not on a par; the requirement of simplicity, closely related to the Machian ideal of conceptual economy, is relegated to the lowest rank. At first blush, it seems to be a purely *aesthetic* criterion, akin to the criterion that mathematicians usually allude to in talking about *beauty* in mathematics. Still, it has a practical significance, especially in the more developed compartments of science, where definite restrictions have to be imposed on the informational complexity of a cognitive representation, if just to account for the finite capacity of humans to process information. The first three requirements pertain more clearly to the fundamental purpose of explication, i.e., the task of providing conceptual tools for cognitive systematization and unification of science. I will comment upon these features in more detail in the course of the discussion of Carus' interpretation of the Carnapian framework of explication. For now, to summarize the general 'structural' conditions of explication, let us take a brief look at the types of concepts that are explicated.

The medium of explication is first and foremost the conceptual space of any particular compartment of knowledge. As we saw above, the process of explication necessarily presupposes an aggregate of more or less informally defined concepts as its data, or *explicanda*. The kinds of concepts that might act as the starting point of an explication can belong to at least the following groups:



The function of concepts could be classified as follows:



The functional classes are described by Mario Bunge in the following terms [Bunge (1967)]:

Formal		
	Basic:	those providing the rational cement, like "and" "all", "set", "square root", "distance" "group"
	Metalogical:	those occurring in analyses and theories of formal theories; e.g., "well-formed formula", "provable", "axiom", "theorem", "theory"
Non-formal		
	Descriptive:	those enabling us to describe experiential matters of fact, like "body", "red", "near", "event", "between", "liquid"
	Interpretive:	those occurring in the interpretations of descriptions, like "species", "atom", "birth rate", "motivation", "inhibition"
	Prescriptive:	those occurring essentially in norms, rules, and conventions — e.g., those designated by verbs in the imperative form.

These schematized presentations illuminate the variety of the 'conceptual engineering tasks' that fall under the purview of Carnapian explication.[5] Moreover, they can be seen

[5]In a more abstract and philosophical level it might be expedient to have recourse to accounts like that of Robert Nozick which aim to systematize the process of structuring of (philosophical) concepts. Nozick's classification is presented in the context of the problem of personal identity but it generalizes to all kinds of

as tools for facilitating the process of clarification that precedes the more exact process of explication. Schemes such as these help in coördinating the different explicative tasks and work as aids in the more general task of a descriptive study of human cognition that underlies Carnap's conception of naturalized epistemology. Carnap's conception, in its broad outlines similar to the one that Quine advocated, constitutes a basis for interpretative possibilities that go far beyond what Carnap ever stated explicitly himself. Building upon this basis, André Carus presents a radical interpretation of Carnap's ideal of explication as an account of reason. I will now turn to his interpretation of Carnap's overall epistemological program.

6.2 Carus' Interpretation of Explication: Possible Vistas for Further Development

André Carus presents his interpretation of Carnap's ideal in his illuminating and meritorious book *Carnap and Twentieth-Century Thought — Explication as Enlightenment* [Carus (2007)]. The starting point of his exposition of explication in Chapter 11 is the observation that the Principle of Tolerance which Carnap introduced in 1932, comprised substantial potential for application which Carnap was unable to fully realize himself: "The full application of the new principle was impeded for many years by lingering prejudices from earlier stages in his development. This delay, combined with Carnap's predilection for working on particular language projects rather than the architectonic of the overall ideal, meant that he never fully spelled out his ideal of explication as an account of reason, and it is left to those following in his footsteps to piece together the hints he left." [Carus (2007), 273] This is precisely what Carus has set out to do. In keeping with the naturalistic framework of Carnap's epistemology, Carus sketches his suggestion for reconstructing the Carnapian ideal by rounding out and supplementing the fragments left by Carnap with pertinent material from the corpus of contemporary scientific knowledge. This sketch intends to convey how the Carnapian ideal "could be relevant to real life", giving it "some social and historical texture." [*ibid.*]

---

concepts and can be utilized in a number of situations:

I. *Intrinsic Abstract Structural.* A concept C's holding at a time is analyzed in terms of an abstract structural description involving only mondadic predicates holding at that time. [The personal identity of something is an intrinsic feature of it, most usefully discussed without considering any entities other than it or any of its features at any other time. (For example, the identity resides in the soul.)]

II. *Relational.* X falls under concept C if X stands in a certain relationship R to another, sometimes earlier, thing of a specified sort. [For example, X is the same thing as the earlier Y if X is spatiotemporally continuous, or psychologically continuous, with Y.]

III. *Closest Relative.* To the relational view is added the condition that nothing else is as closely related under R to that other (previous) thing. [The closest continuer theory of personal identity is of this sort.] [Nozick (1981), 47–48]

6.2.1 The presuppositions of Carus' interpretation

The fundamental presupposition of Carus' interpretation is that Carnap, "faced with the Hegelian task of a history of reason", would have started from a sharp distinction between "two kinds of coding or representation systems". [*ibid.*] Carus refers to these as the "evolved" and "constructed" systems, respectively. This fundamental distinction is known to have been made by Carnap who used such locutions as "informal" and "formal" [languages], or "word-languages" and "artificially constructed symbolic languages", or "natural" and "artificial languages". This distinction is similar to the one, as Carus suggests, made by Gaston Bachelard between scientific and everyday modes of thinking, who referred to it with the emblematic phrase "*coupure épistemologique*" (epistemological discontinuity or rupture). One of Bachelard's original views on rationality and language constitution was the thesis that "we are autonomous and responsible beings (both individually and as a species) only in so far as we construct our conceptual frameworks self-consciously, rather than going complacently with the flow of our inherited concepts." [Carus (2007), 274][6] But despite the rupture, there is also a continuity between the two types of discourse, as Carnap observed more and more clearly, especially towards the late 1930s. Carus notes how Carnap had in the *Syntax* regarded even evolved languages as *calculi*, i.e., essentially as systems of formation and transformation rules.[7] As a result, he was able to conceive of a broad spectrum of different language systems, with one end of the spectrum consisting of languages regarded as unambiguous cases of different types of *calculus*, and the other end consisting of languages the status of which was not as clearly defined, i.e., languages which contained more 'evolved' elements. (All in all, every system contained at least *some* constructed elements.) Carus rejects the suggestion by Thomas Ricketts [Ricketts (2004a), (2004b)] that the artificial languages should be conceived as some kind of conceptual maps or coördinate grids superimposed on a system of evolved language. There is not any natural demarcation line between the evolved and constructed languages, "nor is there any suggestion that a constructed calculus is to be taken as a map or grid constraining a more amorphous evolved language, such as that used by scientists. The constructed languages are to be regarded, rather, as candidate vehicles for the construction of science as a unified theory, including a meta-language in which to construct and evaluate the syntax of the object language (later also its semantics and pragmatics)." [*ibid.*] The historical development of different linguistic frameworks provides an illustration of the mutual interaction between more 'constructed' and evolved systems. For example, as Carus notes, very early in the history of different civilizations there have arisen partly formalized languages devised by groups of specialists to be used as tools in various domains of activity, including the use of specific language systems in government, trade and theological disputation. The evolution of such systems of language has been very gradual. Still, there are ex-

[6] Carus provides here references to [Bachelard (1934), (1938)].

[7] N.B. The classification of *calculi* by their formation and transformation rules has become standard. It is often thought that the distinction was first introduced by Carnap. This is not the case, however. In *Logische Untersuchungen* [1900/1901] Husserl had already made a distinction between the *material (synthetic)* absurdity and *formal, analytic* absurdity (Investigation IV, §14). The rules governing the avoidance of these absurdities correspond to the Carnapian rules of formation and transformation.

amples stemming from Antiquity that demonstrate the *coupure épistemologique* between constructed and evolved languages in a most dramatic way. Especially the development of synthetic geometry in Greece which culminated in the unified system of Euclid with its strictly axiomatic character exemplified the very radical estrangement of the constructive types of language used in science from the language of everyday life. It is solely because of the limited applicability of the language of Euclidean geometry to practical problems of every day life that the rift between these modes of thinking was not experienced as forcefully as we nowadays experience the one between our everyday language and the language of mathematical physics, for example. It is perhaps true, as Carus claims, that "it was only in the seventeenth century, and especially with the publication of Newton's *Principia* in 1687, that it became more widely clear to educated people that the language of genuine knowledge was starkly different from the evolved language of daily life. Only at that point did a scientific culture begin to form, and an international community of those acculturated in the new attitudes and fluent in the new languages. (The existence of such a community, however small and embattled, was the basis for Enlightenment.)" [Carus (2007), 275]

From the point of view of Carnap's naturalized epistemology, it is plausible to allow for the evolution of *both* the evolved and constructed languages. The social scientists informed by a biological or evolutionary point of view regard cultural evolution (of which the evolution of constructed languages is a part) and physical evolution as interacting processes. Especially the competitive aspect of learning and experience-mediated specialization are seen as essential features of social organization in which physical evolution plays a prominent part. This is quite clear in the context where individuals compete for scarce resources of food, for example, but can the evolutionary viewpoint of biology offer anything of real substance in relation to more complex social situations? The evolutionary biologist's view is succinctly put by Mary Jane West-Eberhard:

> Learning takes an entirely new aspect when competition is for social status and socially defined rewards, rather than other kinds of resources, especially in organisms, such as humans, and perhaps other animals, capable of manipulating and learning socially defined rewards and punishments. Then social development (socialization) can come to occur within a framework where successful behaviors change from group to group, and generation to generation, for success no longer is defined by some fixed or slowly evolving criterion, such as ability to find and handle particular kinds of food (which might be sufficiently recurrent to produce genetic correlations between phenotypic traits). Under socially defined competition, success has relatively flexible and changeable criteria invented and imposed by parents and influential dominant leaders, reinforced by mimicry and strong rewards and punishments that enforce both conformity and change. This is one reason why it can be extremely misleading to suppose that particular human learned behavioral traits are evolved. The role of genetic variation and evolution should be sought in any traits that would recurrently influence success despite rapid social change, such as generalized ability to communicate, observe and adapt

> to shifting criteria of success, and so forth. This follows from the principle that to be molded by natural selection, a trait not only must be heritable (show genetic variation correlated with phenotypic variation) and affect reproductive success, but also must be sufficiently recurrent in a population across generations for natural selection to act. [West-Eberhard (2003), 350]

The Carnapian viewpoint is not completely subsumed under the explanatory framework of biology, however. As Carus remarks, "this empirical viewpoint is compatible with an *ideal* viewpoint that places positive value on the one kind of evolutionary process because of its role in liberating the human species (especially those who seek it out) from unreflective passivity in thought and action." [*ibid.*, 275–276] The core tenet of Carus' interpretation is that "[f]rom Carnap's ideal viewpoint, [...] the liberation of human thought from passive complacency and the shackles of the past depends on the progressive replacement of evolved by constructed languages." [*ibid.*][8] It is easy to agree with Carus' thesis that explication aims less to *describe* the relation between two kinds of linguistic system in social practice than to *raise* the degree of construction (and precision) in human languages. This, as we saw, was already emphasized by Carnap in Chapter 1 of *LFP*. What is not as fully spelled out in that chapter is the problematic associated with the thesis that theory choice and language choice are to be regarded as practical matters. In the level of external questions about the applicability and fruitfulness of a certain proposal (for a linguistic system to be used for a definite purpose), the sentences used to evaluate the options no longer have "cognitive" meaning; indeed, some must have normative or "optative" meaning.[9] The thesis that the ideal of explication determines an area of normative pragmatics, i.e., a framework for improving our systematic pursuit of knowledge in the way of elaborating devices for this purpose — e.g. various communicative systems — needs to be scrutinized from precisely this 'optative' viewpoint. Are there explicitly spelled out optative rules that determine such a normative pragmatics? Let us scrutinize Carus' answer to this particular problem in more detail.

6.2.2 Criticism of Carus' position

After having spelled out the Carnapian requirements for explication, Carus writes: "[T]he choice among alternative explications is a practical problem. It must be made

[8] Carus cites Carnap [ASP (1956a)] to back up the thesis about the importance of the constructed languages in Carnap's epistemology: "I do not share the apparently widespread view that the vagueness and ambiguity of most words in everyday language do not much interfere with human communication. It is hard for me to understand how someone could really believe this, in view of the countless misunderstandings and failures to get something across that we observe daily. I would have thought that there could be no disagreement about the damage done by this vagueness." [*ibid.*]

[9] This is a locution which Carnap uses in the section VI titled "Value Judgements" in [Carnap (1963), 999–1016]. There Carnap gives the following characterization of "optative": "There is a general kind of meaning common to all statements expressing a wish, a proposal, a request, a demand, a command, a prohibition, a permission, a will, a decision, an approval, a disapproval, a preference, or the like, whether or not they also contain meaning components referring to matters of fact. I shall call a sentence which, among others, has meaning component of this kind an 'optative sentence' or for short 'an optative'." [*ibid.*, 1001]

in the context of action, which overlaps to some degree with the *Lebenswelt* in which the participants articulate the values and preferences that guide their choices. The *process* of choice can of course itself be formally represented within a decision theory or game theory or other formalism, but the values and preferences embodied in the utility functions that occur in the relevant interpretations of such formalisms must still be supplied from outside by the people making the choice." [Carus (2007), 280] Yes, indeed. But what must be made explicit is the actual *process* by which the values are arrived at, and precisely *how* these values coördinate the regulative constructed language systems. This is an issue to which Carus provides a surprisingly insufficient answer, or, regarding the category shifting he undertakes, no answer at all. Elaborating on the issue of the value background of such language engineering projects, he says: "For Carnap, then, the context of action and the context of knowledge are not mutually interreducible. He would of course agree with Quine (who also acquiesces in an is-ought distinction [...]) that our values are just as much part of nature as our organisms or our social interactions or whatever else produces those values [...], and that the theory of nature is thus also a (descriptive) theory of values." [*ibid.*] This seems to me an unsatisfactory line of argument. In the first place, the dynamic interplay between the language construction process and the values that guide this process is not made any clearer by relegating the problem to the level of (routine) science in the hope that, given sufficient time, it will discover the mechanism of "social interactions or whatever else produces those values". In the second place, although such a solution might be (hypothetically) found, it does not relieve us from the responsibility of carefully scrutinizing the presuppositions of such a hypothesis. Although the thesis about the reducibility of the domain of values to the domain of facts might in itself be philosophically insulting to some, the far more serious neglect is Carus' utter and disastrous silence about the relationship between systems of regulative constructive language and the coördinating function of values in choosing a particular linguistic framework. This lapse severely undermines the plausibility of the thesis that his reconstruction of Carnap's framework vindicates the Enlightenment core of Carnap's philosophical program.

Carnap's program of explication, as we have seen, is clearly undergirded by the Principle of (*logico-mathematical*) Tolerance presented in the *Logische Syntax der Sprache* (published in 1934). The importance of the liberation brought about this transition within the domain of knowledge acquisition and systematization can hardly be exaggerated. Carnap's notion of explication that opens up a whole new space of possibilities (in a truly Musilian spirit of a *Möglichkeitswissenschaft*) for language construction and scientific communication, must be seen as a potential vehicle for bringing order to the occasionally frustrating juxtapositions between different proposals for a theory in a particular domain. Exemplary domains of scientific activity where a Carnapian approach might bring about new possibilities for scientific dialogue are easy to find. We have mentioned several examples in the preceding sections, such as open problems within the foundations of physics, philosophy of science and historiography.

I do not want to imply that Carus' reconstruction of Carnap's ideal is in the wrong track. On the contrary, I think he has captured the essential core of Carnap's covert philosophical motivation. What I find wanting in the reconstruction is the casual by-passing of the value-theoretical background that underlies Carnap's program. True, Carnap never spelled out the ethical presuppositions of his program, although there are definite hints that such a background framework was really active in his thought (this is exemplified in his correspondence with Neurath, for example). There are some important remarks about the relationship between beliefs and attitudes (which is precisely what the Principle of Tolerance is about) and about the compatibility of difference in attitudes with agreement in all (relevant) beliefs in the Section VI: "Value judgements" in [Carnap (1963)], but on the whole the discussion focuses only on the explication of certain classes of optative statements without tapping into the moral presuppositions of those statements.[10] Although a superficial assessment might guide us to believe that Carnap advocates a hollow moral pluralism, there is one conspicuous feature in his philosophy that talks back against this: the Principle of Tolerance. This principle is calculated to guarantee the gradual improvement and precisification of the linguistic tools that we use to describe the world. Moreover, it is calculated to guide the process of evaluation and assessment of the different proposals for such linguistic tools in a way that leads to a *convergence* of the representations that we devise and utilize to describe the world. The unity of science (and knowledge more generally) was an ideal that Carnap never abandoned. Precisely because this ideal undergirds much of Carnap's work in philosophy, it would be essential to scrutinize more carefully the presuppositions of this ideal and to spell out clearly the moral framework in which the Principle of Tolerance is embedded. This is an issue that needs to be delved into in far greater detail than is possible here. My intention has been only to show that, if Carnap's Principle of Tolerance is to be taken seriously, a reconstruction of the value-theoretic background of his ideal cannot be eschewed. Carus does not take up this task, as we have seen. There is, however, a very important aspect to his representation of the Carnapian ideal which I wholeheartedly endorse. I will return to this aspect in the following sections. But on Carus' reading, what is left of the Enlightenment core of Carnap's ideal, deprived of the moral basis, is bound to remain an impression of an ethereal azure:

> The ideal does not claim to know what reason *is*. It imposes a minimal constraint — any larger conception of reason must accommodate our knowledge; it cannot be obviously inconsistent with what we know. But our knowledge is

[10] The conclusion of the "Value Judgements" hints at the fact that Carnap might have been interested in spelling out his moral philosophy in more detail: "In line with my general tendencies, I would regard it as advisable to propose explications of value statements in standardized forms in a constructed language. I think the fact that standardized forms for cognitive statements, both logical and factual, have existed for a long time, has contributed significantly to our clearer understanding of the nature of these statements and the logical relations between them. I think it will hardly be possible to state precise rules for translating value statements of the ordinary language, including those of the customary philosophical language, into standardize forms, because the customary forms are too vague or ambiguous. The same hold of course also for the customary cognitive statements. However, it is important that the meanings of the standardized value statements should be made sufficiently clear, at least for practical purposes, as has been done for cognitive statements. Furthermore, logical rules must be stated for the logical relations, especially for logical implication, both between value statements and between value statements and cognitive statements. [...] If I had time to devote myself to this task, I would try to develop explications in the direction indicated,on the basis of the analyses which have been given so far by pragmatists and logical empiricists. The direction of my own work would probably be closest to that of Stevenson and Reichenbach." [Carnap (1963), 1012–1013]

> so limited, compared to the vastness of our ignorance, that this is a very loose constraint. It is much too early in the history of systematic human cognition to make pronouncements about the ultimate nature of reason. We are in our cognitive infancy. Sixty years after he first set his sights on the open sea of free possibilities, it still lies before us, all but unexplored. We have been extremely timid, clinging to the shoreline, hardly daring to venture out of sight of land. The warm, familiar, safe harbour of habit and tradition appeals to us as much as it ever did to our ancestors. It is time we ventured forth again in the pioneering spirit of the Original Enlightenment, emboldened by Carnap's example. [Carus (2007), 309]

6.3 History of Particular Explications as a Prerequisite of Explication

As Carus remarks in his book, "Carnap never got very far in characterising the kind of external discourse in which 'theoretical investigations and practical deliberations and decisions with respect to an acceptance or a change of frameworks' would be carried on. He gave it a name, 'pragmatics', but remained vague about the details." [Carus (2007), 266] That Carnap envisaged such an area of scientific discourse is interesting. The idea of such a framework for addressing external questions was, of course, in stark contrast with the position of Quine for whom all questions were internal. Quine famously regarded the supposed distinction between semantics and pragmatics a "pernicious error." [Quine (1987), 211] Carnap, although entertaining hopes of making pragmatics a technical subject in keeping with his philosophical predilections, remained quite realistic about the prospects of constructing a formal theory of pragmatics. To provide an example, Carus relates how Carnap's student R. M. Martin had given a proposal for a technical apparatus for pragmatics [Martin (1959)], but that Carnap had taken a reserved stance, writing to Evert Beth, the editor of the series to which Martin's book had been submitted, that "it might have been best to clear ground first with more preliminary clarifications, as agreement had not yet been reached even there." [*ibid.*] Carnap elaborated the general idea further:

> Since probably at the present time various authors would choose different concepts and different methods, it might at first be advisable to write a less technical treatise explaining the concepts, defining them informally in such a way as to indicate clearly how they would be defined in a formalised metalanguage without necessarily giving actually their formalised definition, studying alternative explications of the various concepts, etc. (What I have here in mind is roughly analogous to Russell's preparatory discussions in his *Principles of Mathematics* before writing the *Principia* or my preliminary discussions in *Meaning and Necessity* in preparation for a not yet written treatise of a semantics of extensions and intensions in either an extensional or an intensional

formalised metalanguage.) [ASP (1958a)][11]

Comparing the tenets of Carnap and Quine, it is easy to overlook the issues over which they were largely in agreement. Indeed, regarding both as representatives of "naturalistic epistemology" (a locution used by Quine), the convergence of their views on one aspect of epistemology is notable. Although they classified the activities pertaining to knowledge acquisition differently, there was a purely scientific, descriptive element to it that brought Carnap's view closer to that of Quine. As Carus emphasizes, "this was a part of what he called descriptive pragmatics — the descriptive study of human cognition in all its aspects as an essential basis for the practical task of conceptual engineering. It was essential for that task because the biological, physiological, psychological, linguistic, sociological, economic, and historical knowledge about our cognitive capacities and the development of our actual knowledge are the data we have at our disposal when we seek to construct languages and decide among explications. They define the existing possibility space within which we construct languages adequate to what we know already, and from which we can extrapolate to imagine new possibilities." [Carus (2007), 267] Carus' view on the status of the descriptive element in the method of explication is succinctly presented by the following analogy:

$$\frac{\text{descriptive study of human cognition}}{\text{explication}} = \frac{\text{biological knowledge}}{\text{medicine}} = \frac{\text{physical knowledge}}{\text{structural engineering}}$$

The relevance of the descriptive study of human cognition for the Carnapian project of explication is, according to Carus, buttressed by the fact that "naturalistic epistemology in this broad sense has made great strides since Carnap's death, on many different fronts. We know much more now about the processing involved in the various human sensory systems, and can model higher-level capabilities like understanding language and written texts. One of the early protagonists of the cognitive revolution, Noam Chomsky, had used Carnap's own idea of regarding natural languages as systems of rules. This rule-based approach has been challenged by theories that modeled human processing on self-organising weighted networks (parallel or 'connectionist' processing). Hybrids between these two approaches, seeking to portray human thought as a dialectical interplay between the two poles of receptive processing and active shaping of a 'situation model' [...] recapitulate a basic Kantian motif and are thus reminiscent of Helmholtz, who has justly been called the father of cognitive science." [*ibid.*, 268] The accelerated progress in these fields has been amazing, of course. What seems to pose a significant challenge to scholars working in this vein is the social dimension of a "naturalized epistemology" that rests upon the detailed findings of cognitive science. The situation is especially problematic in connection with the constitution of scientific knowledge and the basis of scientific rationality. How exactly is the social dimension of scientific activity accommodated in the viewpoint of cognitive science? What are the specific social *mechanisms* responsible for the process of knowledge acquisition and systematization that takes place under the

[11] Cited in [Carus (2007), 267n]

purview of the scientific enterprise? What is often taken as a critical shortcoming of Carnap's approach is, as Carus rightly remarks, that "he did not in practice distinguish between the scientific enterprise [as a sociological phenomenon] and the knowledge generated by it as an objective artefact." For us this distinction has become something of a *habitude* of thought, thoroughly infused as we are with the ideas of the tradition of science studies that took shape in the wake of Thomas Kuhn's hugely influential *Structure of Scientific Revolutions*. What Kuhn perpetuated was a systematic application of empirical inquiry in studying empirical inquiry. When Carnap (who was the editor of the series in which Kuhn's book appeared) encountered Kuhn's book, he embraced it with enthusiasm. In fact, he regarded it as both a natural and important supplementation of his own work in inductive logic. [Reisch (1991)] But it is doubtful that he would have approved of the general view on the scientific enterprise that the 'Science Studies' had constructed by the end of the twentieth century. He did not foresee this development of course, the "Pandora's Box that Kuhn's book would open". [*ibid.*, 269] As Carus puts it, "[h]e could not have conceived of the situation that has now resulted, in which there is no agreement among historians and philosophers of science even on the most basic assumptions about the character of the enterprise." What is most needed, then, is "a pluralistic view that regards the different positions as proposals for the form of meta-language for the study of science." Only such a view "could have any hope of bringing about whatever mutual understanding might still be possible." [*ibid.*] It is precisely here that Carus' interpretation of Carnap's program proves to be most illuminating. Taking into account the highly perplexing situation in philosophy and history of science, Carus proposes that Carnap's method of explication, the prerequisite of which, as we saw above, is a process of clarification, suggests "a program for clarificatory pragmatics consisting in a certain kind of conceptual history of the scientific enterprise, with special attention to episodes in which major changes of theoretical direction were undertaken." [*ibid.*] Such 'internal' histories should then be supplemented with the point of view provided by *external histories* of science, including sociology, psychology, economics, and anthropology of science, which all work in the service of descriptive pragmatics. But this is not an adequate picture of the whole range of historically informed studies of science, as Carus readily emphasizes. What is, in my view, the most illuminating aspect of his interpretative framework is the importance he places on histories of *particular explications*:

> But even the sum of such internal and external histories do not yet suffice to yield what Carnapian pragmatics actually requires to support the program of explication, the external discourse in which 'theoretical investigations and practical deliberations and decisions with respect to an acceptance or a change of frameworks' are pursued. For this we need, rather, something that can shed light on the history of explication itself, or more concretely, the history of particular explications. It is only against the background of such histories that the *present* activity of explication can be properly informed about its limits and its possibilities. [Carus (2007), 270]

This viewpoint is in perfect harmony with the Machian conception of history of science as a background framework for our creative tasks of knowledge acquisition and systematization which I alluded to in Chapter 1. If we believe Mach, history of science is irreplaceable; no other discipline could equally act in the role of facilitating the process of devising new and fruitful theories and conceptual tools in the way of opening our minds to conceptual possibilities on the one hand, and disclosing tacit principles and modes of thought on the other. The project of historical elucidation of the different particular explications in a given field is thus seen to be one of the most important undertakings in philosophy of science. In keeping with this observation Howard Stein has drawn a distinction between "the enterprise of knowledge" and the "enterprise of understanding" [Stein (2004)], thus showing us how acutely conscious he is of the fact that to be able to *use* our knowledge, we must understand it (at least to some degree). The "enterprise of understanding", then, as Carus remarks, "at least partly *consists* in constructing and explaining precisely those conceptual genealogies, sequences of explications, that provide present tasks of explication with their possibility space." [Carus (2007), 270] Such an approach might well be called, according to Carus, *philosophical history of science*. I will now proceed to give an example of the problems that this kind of history of science could tackle with.

6.3.1 An example of *philosophical history of science*

Probabilistic concepts in physics

In this section I will illustrate with an example how the prerequisite of explication, viz., the task of providing *genealogies* of important scientific concepts and ideas, proceeds in the intersection of mathematics, physics and philosophy. I concentrate on the application of probabilistic notions in physics and the perennial problem of determinism. Giving an account from a historical point of view of one of the most recalcitrant problems in mathematics and physics, the *three-body problem*, I illustrate how the approaches of all these different disciplines supplement each other and contribute towards acquiring an enhanced understanding of one of the most challenging issues in modern science. I begin with a general note on the significance of probabilistic notions in physics.

The interrelation between the development of physics and of probability theory is from a philosophical point of view a relatively little researched, but intellectually stimulating topic. Probabilistic notions were first developed independently without recourse to practical problems in physics. As is well known, this took place in France in the 1650s, when gambling was a popular and fashionable habit. A passionate gambler, the chevalier De Méré, had the idea of consulting one of the most illustrious amateur-mathematicians of his time, Blaise Pascal, on some questions connected with certain games of chance. This instigated Pascal to begin correspondence with some of his mathematical friends, most notably Pierre de Fermat. This correspondence marked the origin of classical probability theory. But it was only after the development of classical statistical mechanics in the late 1800s that the probabilistic concepts started their invasion to physics. This had important consequences for the development of modern probability theory. Another impulse came from astronomical dynamics where the problem of determining the distribution of the partial quotients in the continued fraction expansion of a real number was crucial for the application of perturbation theory. The latter problem, which had seemingly no connection with more general problems in physics, led in the hands of Henri Poincaré to the first sketch of the theory of dynamical systems. Moreover, the computational possibilities for applying these notions widened considerably during the latter half of 20th century as modern computers became available. This opened up a completely new area of research, the theory of computability, which brought about a number of interdisciplinary innovations in mathematical logic, computer science, mathematics, physics and even biology. The most important new ideas in the theory of computability were provided by Kurt Gödel, Alonzo Church, Stephen Kleene and Alan Turing, among others. In addition to these developments the scientific community witnessed the birth of the fundamentally indeterministic physical picture of quantum mechanics in the 1920s, which received its first coherent formulation in Heisenberg's matrix mechanics of 1925. As von Plato relates in his historical study on the foundations of modern probability theory: "The concept of probability in quantum mechanics arose through Born's work in the middle 1926, from the soil of Heisenberg's and Einstein's transition probabilities, de Broglie's matter waves, Einstein's gas theory and Schrödinger's wave function." [von Plato (1994), 17] It is arguably true that without the changes brought about the development of modern physics, modern probability theory could not have obtained the remarkable position it today enjoys in scientific and philosophical thinking.

This, however, is not the whole story. Firstly, the transition from classical to modern probability is not at all as well documented as one would be inclined to think. Jan von Plato's monograph *Creating modern probability* [*ibid.*] is one of the rare contributions towards clarifying the conceptual developments in the early 1900s in probability theory. More work remains to be done. Secondly, a thorough historical and philosophical analysis of the importance of computational ideas in the attempts to intellectually grasp the nature of some intricate dynamical phenomena in physics and astronomy — e.g. the three-body problem — is wanting. The history and evolution of the problems that gave rise to the theory of dynamical systems and their connection to the computational ideas, such as the theory of complexity, remains inadequately charted. This is an especially grave shortcoming in the light of the attempts to found an objective interpretation of probability on the basis of ergodic theory. The task of clarifying and representing the historical evolution of these seemingly disparate areas of research will in a very important way open new conceptual possibilities for the clarification of the philosophical problem which concerns probability and its role in physics. Thirdly, the significance of the notion of probability in quantum mechanics is still widely disputed. On the one hand it is contested which fundamental principles the probability calculus of QM is to be founded on. If the program of ergodic theory is entertained, a suitable candidate for the foundations of quantum mechanical probabilities is the quantum stochastic calculus developed among others by K. R. Parthasarathy. On the other hand some interpretations of quantum mechanics make explicit use of probabilities in the epistemic sense, that is, they are inherently subjectivist in nature. How exactly this is going to be reconciled with the objectively interpreted probability amplitudes predicted by Schrödinger's equation remains an open problem.[12]

[12]In my view, the attempts of David Deutsch, David Wallace and Hilary Greaves to reformulate the Ev-

I think that here as well, Carnap's method of explication finds its most natural field of application. A much broader and deeper account of the theoretical and experimental details of modern quantum theory is called for, and such an undertaking is ideally suited to researchers acting in several interdependent fields.

One of the most prominent areas where explication might prove its mettle is the very interesting current attempt to devise a coherent theory of Quantum Gravity. The discussions within this area essentially center around the question about which form the putative fundamental theory of physics should take. Given the cornucopia of different approaches in quantum gravity (string theory, loop quantum gravity, QFT in curved spaces, lattice approaches, Euclidean quantum gravity, non-commutative geometry, quantum cosmology, twistors, etc.) there is a genuine need for a general framework for evaluating the different proposals in a non-biased manner. It seems to me that this situation in physics presents a proving ground *par excellence* for the kind of conceptual engineering tasks coördinated by the ideal of explication that Carnap envisioned in the twentieth century. The multiplicity of the mathematical models and the variety of different theoretical frameworks flourishing in this particular field is primarily an indication of the fact that progress cannot be obtained on a purely formal or technical level, but that we are essentially dealing with a very complex problem in the determination of which all the following ingredients enter in a fascinating way: rigorous mathematics and bold speculations, concrete physical models and general schemes, fundamental questions and technical issues as well as phenomenological scenarios (which refer to models which are not explicitly based on derivations from fundamental laws but operate on an 'emergent' level). The conceptual problems are mainly semantical: it is not entirely clear how the different mathematical proposals are related on the level of fundamental physical concepts and theory construction. One of the major watersheds among the theory-candidates is the assumption of a background space-time: some theories (such as string theory and M-theory) presuppose the existence of a background space analogously with quantum field theories, whereas others presuppose that the quantum states themselves define and determine the space-time analogously with the way in which the solutions of the equations of General Relativity determine space-time. A clear account of the general situation is provided by Carlo Rovelli, one of the leading researchers in the discipline: "Conceptually, the key question is whether or not it is logically possible to understand the world in the absence of fundamental notions of time and time evolution, and whether or not this is consistent with our experience of the world. [...] The difficulties of Quantum Gravity are indeed largely conceptual. Progress in Quantum Gravity cannot be just technical. The search for a quantum theory of gravity raises once more old questions such as : What is space? What is time? What is the meaning of 'moving'? Is motion to be defined with respect to objects or with respect to space? And also: What is causality? What is the role of the observer in physics?" [Rovelli (2009), 8; in *Approaches to Quantum Gravity* (2009), Cambridge University Press]

Since its honorable beginnings physics has evolved to an extraordinarily rich discipline with a wealth of subdisciplines. This has also meant increasing interaction between

erettian interpretation of QM by means of decision theoretic concepts seems quite problematic.

physics and computer science. In the first place, in a practical level, computer science has given physics mainly computational power, creating within physics whole new areas of specific interest such as soliton theory, chaotic dynamics and quantum lattice theory. But, in the second place it has given rise to an entirely new conceptual approach to physics. In a sense, it has brought about a diffusion of concepts, one of the most extreme examples of which is the idea of the physical world as a computer. It is these more fundamental, ontological ideas, that have proved to be the most important ingredients in the discussion about the importance of computability theory in physics. However, it seems that the more sober approach of tracking the historical development of the basic ideas of computability and their relation to probabilistic notions in physics would also shed considerable light on the more fundamental questions about the computable features of physical processes. For this purpose it would be expedient to assess to what extent the classical dynamical problems in physics and astronomy have worked their way back to the core of the discussions about the role of randomness in nature and probabilistic laws in the physical sciences. Indeed, notwithstanding the intriguing questions brought up by the recent developments in quantum mechanics concerning the prospects for quantum computation and the algorithmic nature of all physical processes, it can still be argued that our understanding of probability and probabilistic laws in physics will be enhanced by familiarity with the classical dynamical problems in physics and astronomy. I will briefly adduce a putative line of research with respect to these questions in the following section.

On determinism, randomness and chaos in classical mechanics

As mathematicians, physicists and philosophers of science well know, the fascinating interplay between the notions of determinism, randomness and chaos constitutes one of the most profound conceptual puzzles of modern science. It is my goal in this section to shed some light on the origins of the theoretical developments that have led to the observation that essential to an understanding of determinism in a more philosophical sense is an appreciation of the varieties of ways it works in physics. It is often said that Newtonian physics presents the paradigmatic example of causal determinism. Nothing could be further from the truth, however. In this section I reiterate the reasons for this, drawing on the accounts of determinism of John Earman and Patrick Suppes. My exposition is far more limited, though, than the one provided by [Earman (1986)], for example. I attempt to present only some of the important lines of the historical development that have led to the appreciation of the richness of the notion of determinism in physics. I focus explicitly on the three-body problem and its significance for the development of the contemporary notions of determinism and instability.

The most prominent contribution in this area has been made by Henri Poincaré whose pioneering researches in celestial mechanics opened up the way for the theory of dynamical systems. Poincaré is the first mathematician to have illustrated the ideas pertaining to what is known today as *chaos*. Another important line of research, not altogether unrelated to Poincaré's, is the theory of random sequences first articulated by the Russian mathematician A. N. Kolmogorov, and subsequently elaborated by Gregory Chaitin, Per Martin-Löf, and Michiel van Lambalgen. I attempt to show that these seemingly disparate lines of mathematical research come together under the purview of the theory of dynamical systems and that the interaction between them has resulted in the most efficient conceptual tools for analyzing the role and meaning of determinism in modern science. At the very least, I hope that I can convey the reader the necessity of thoroughly examining the workings of determinism, randomness and chaos in physics to get a firm grasp of their conceptual relations. A proper place to start is classical physics, more particularly celestial mechanics, where all of these notions come to play a significant role.

Preliminary remarks on determinism

To begin with, we need a decent enough working definition of determinism. But this poses us a challenge of its own. As John Earman has made clear, "we cannot begin to discuss the implications of physics for the truth of the doctrine of determinism until we know what determinism is; on the other hand, no precise definition can be fashioned without making substantive assumptions about the nature of physical reality [. . .]". [Earman (1986), 4] Any working definition (of determinism) is bound to be somewhat vague. In Earman's impish phrase: "the price for this vagueness is the loss of precision or generality, or both." [*ibid.*] But this is not fatal at this point: as we investigate a concrete example from the history of celestial mechanics, the boundaries of the notion become sharper, and hopefully we begin to get a firmer grasp of what is at issue.

The following working definition of determinism is quite familiar. However, it is not the only intuitive conception of determinism in the market (even household determinism comes in many flavors). What I wish to convey is the extraordinary difficulty arising from the complex interplay between epistemological and ontological issues. Precisely because of the rough interface between these domains in the problem of determinism it is of utmost importance to clarify both aspects of the problem; undue emphasis on either aspect results in a deficient conception of determinism. Philosophically it would be desirable to be able to answer the following questions: What are the main *conceptual constituents* of the notion of determinism? What are the *problems* that it allows us to define more accurately and how does it confine the *possibilities for solution* of these problems? I do not intend to answer these questions in this section; I will merely cover some background both historically and conceptually that would facilitate the work on these general, yet delicate, problems. The areas that are essential to understanding determinism include at least traditional philosophy of science, mathematics, metaphysics, practical physics and history of science. The pluralistic approach is called for by the very nature of the problem; only by interweaving these disparate threads of erudition shall we be able to delve into the intricacies of determinism.

A working definition for determinism

The definition that shall work as a starting point for our inquiry has a connection to a quite artistic vision of William James, who in a 1884 lecture to the Harvard Divinity School stated:

> What does determinism profess? It professes that those parts of the universe already laid down absolutely appoint and decree what the other parts shall be. The future has no ambiguous possibilities hidden in its womb: the part we call the present is compatible with only one totality. Any other future complement than the one fixed from eternity is impossible. The whole is in each and every part, and welds it with the rest into an absolute unity, an iron block, in which there can be no equivocation or shadow of turning. [James (1956), 150]

This vision encapsulates the gist of what John Earman [Earman (1986), 5] has called the classical world picture. It consists of the view that the spatio-temporal structure of the world is assumed to embody an absolute or observer-independent simultaneity. Thus, contrary to the lesson we allegedly ought to have learned from special relativity, 'the-world-at-a-given-time' is a meaningful concept. Furthermore, at each instant of time, the world is completely characterized by specifying the values of all the relevant (physical) magnitudes namely instantaneous positions and velocities of particles, instantaneous values of electric and magnetic field vectors, gravitational fields etc. Nevertheless, this form of determinism does not assume materialism or mechanism in any narrow sense. All that is required is that all relevant variables or magnitudes have a spatio-temporal representation.[13]The definition of determinism that is probably quite familiar and the one which I deem adequate for the present purposes is this:

Classical Determinism The world is governed by (or is under the sway of) determinism if and only if, given a specified way things are at a time t, the way things go thereafter is fixed as a matter of natural law. [SEP: "Causal Determinism"[14] [Hoefer (2010)]]

The above definition is often alternatively taken to present the definition of *causal determinism* (a purely terminological difference). This is likely to lead to a confusion. Thus, a few things need to be emphasized. I am explicitly confining the issue of causation outside the present discussion. There is a regrettable terminological confusion pertaining to the use of the terms causation and causality in the philosophical literature. On the one hand, I will here stipulate causality to mean the phenomenon of determinism as captured by the above definition. On the other hand, causation is a concept that I take to

[13]This requirement being compatible, for example, with the Leibnizian ontology of monads, which by my lights cannot be rendered materialistic (straightforwardly), no matter which way you look at it.

[14]The acronym "SEP" refers to the internet portal *Stanford Encyclopedia of Philosophy*. This particular reference can be identified as `http://plato.stanford.edu/entries/determinism-causal/`.

pertain especially to the notions of cause and effect, and the plethora of problems relating to them. These problems are notoriously obscure.[15] Therefore the strategy of tackling determinism via causation seems to me impetuous. We will be better served by trying another route. Maybe causation could be approached more fruitfully once we have a firmer grasp of determinism *simpliciter*,[16] but for our present purposes such a detailed examination is not needed. I will therefore leave this issue aside. Let me now proceed to introduce definitions of determinism that put the emphasis differently from the distinctly ontological mode presented above.

Laplacian determinism

For nearly two hundred years an alternative vision of determinism has greatly affected the thoughts of eminent scientists and philosophers. For some this approach has had considerable appeal. It was first formulated eloquently by Pierre Simon de Laplace, whose work has been pivotal to the development of mathematical astronomy. However, the view in question is not to be found in his *magnum opus* on celestial mechanics,[17] but in the equally awe-inspiring book on probability theory [Laplace (1820)]. In it one finds the following passage in which a metaphysical tone is quickly modified to an epistemological one, conflating the nuances of Jamesian determinism with the scheme of predictability:

> We ought then to regard the present state of the universe as the effect of its anterior state and as the cause of the one which is to follow. Given for one instant an intelligence which would comprehend all the forces by which nature is animated and the respective situation of the beings who compose it — an intelligence sufficiently vast to submit these data to analysis — it would

[15]This at least is the case with the majority of philosophical literature dealing with the subject. I think it is yet premature to try to assess how the concepts of determinism and causation would interact given the theoretical framework of Judea Pearl [Pearl (2000)] for explicating causation (Pearl's book is from my point of view misleadingly titled *Causality*!). Pearl is mainly working with probabilistic models, whereas in this essay I am explicitly concerned with (systems of) deterministic differential equations. The issues are not unrelated, however.

[16]D. H. Mellor would certainly disagree, for he has chosen the exactly opposite strategy of tackling causation first [Mellor (1995)]. Only afterwards has he something substantial to say about physics and determinism, but this seems to me to be a dubious order of procedure. The main problem with Mellor's approach is, I think, that it is confined explicitly within the boundaries of metaphysics. For he says: "Some philosophers reduce their metaphysics to physics, others to logic and semantics. I reduce mine to neither, while taking account of both. So while I accommodate the relevant results of modern physics, I will not for example leave it to quantum physics to tell me whether causation can act immediately across space-like intervals. On the contrary: only when our metaphysics has told us what causation is can we see if physics could reveal unmediated action at a distance (it couldn't)."[*ibid.*, 5.] I am not able to see how Mellor is taking the relevant results of physics (or logic and semantics for that matter) into account, if it is only metaphysics that can provide the answer to the question what causation is. What I feel necessary to emphasize is the pressing need of studying the most intricate conceptual tools of philosophy of science in conjunction with the concrete problems provided by natural sciences.

[17]Laplace's results pertaining to celestial mechanics are compiled in the five-volume Mécanique Céleste (1799-1825).

> embrace in the same formula the movements of the greatest bodies of the universe and those of the lightest atom; for it, nothing would be uncertain and the future, as the past, would be present to its eyes. The human mind offers, in the perfection which it has been able to give to astronomy, a feeble idea of this intelligence. Its discoveries in mechanics and geometry added to that of universal gravity, have enabled it to comprehend in the same analytical expressions the past and future states of the system of the world. Applying the same method to some other objects of its knowledge, it has succeeded in referring to general laws observed phenomena and in foreseeing those which given circumstances ought to produce. All these efforts in the search for truth tend to lead it back continually to the vast intelligence which we have just mentioned, but from which it will always remain infinitely removed. This tendency, peculiar to the human race, is that which renders it superior to animals; and their progress in this respect distinguishes nations and ages and constitutes their true glory. [*ibid.*][18]

This is a portentous view indeed. Although it has much of the aesthetic charm of James's rendering of determinism, it lacks definiteness. If this approach to determinism is to work, we first need a standard (an ideal epistemic agent analogous to Laplace's demon) for assessing which phenomena lie in principle within the sphere of predictive power. But this seems somehow obscure. How is this standard going to be defined? What are the characteristics of an ideal epistemic agent? Furthermore, how are we going to accommodate the progressive nature of science within the scheme of a fixed epistemic viewpoint? That is, how do we guarantee that the standards of the ideal epistemic agent actually render it capable of accounting for all of the possible configurations of natural phenomena, a majority of which certainly lie beyond our present capacities for both mental and theoretical presentation? In fact, these standards are bound to be merely arbitrary rules to settle which natural processes we regard as deterministic (mechanistic, computable, etc.). Even if this definition were feasible, it would be bound to change with time; an awkward result given that one of the intuitive ideas behind determinism is that of immutability (cf. James's vision above). Even though our whole project is about assessing different senses of determinism, the epistemic approach appears to provide too many of them. For example, the most efficient Cray computer, a hypothetically more efficient quantum computing device, or a Turing machine would all yield different senses of determinism. How to settle this question? How to arrive at a satisfactory definition of determinism-as-predictability if the notion of predictability depends essentially on the attributes and skills of the agent that makes the prediction? Perhaps the most cogent way is to drop the condition for predictability out of hand and concentrate on the ontological features of determinism. Be that as it may, the epistemic approach has taken such a firm grip of some notable philosophers of science, that we must provide a glimpse of their views and (mis)conceptions about determinism. To these I now turn.

[18] P. Suppes' translation in [Suppes (1974), 5]

Popperian determinism

Karl Popper distinguished two senses of determinism. He referred to these as *metaphysical* and *scientific* determinism, respectively. Now, what did Popper mean by the former? I do not see a better way of explicating it than the vivid rendering of James quoted above. Therefore, it seems to me that with the term "metaphysical determinism" Popper alludes rather to a tendentious presupposition of determinism than to a genuine doctrine concerning it. A realist in spirit, his intellectual inclinations led him to explicate determinism (as a philosophical doctrine) from a starting point quite different from the Jamesian view. The latter variant of determinism, which Popper called scientific, seemed to him adequate as a philosophical definition of determinism. Popper thought that the doctrine of scientific determinism concerned only that part of the world that was described by a scientific theory. To be more precise, according to Popper scientific determinism is

> the doctrine that the state of any closed physical system at any given future instant of time can be predicted, even from within the system, with any specified degree of precision, by deducing the prediction from theories, in conjunction with initial conditions whose required degree of precision can always be calculated (in accordance with the principle of accountability) if the prediction task is given. [Popper (1982), 36]

The connection with determinism of the Laplacian variety is obvious. But what distinguishes Popper's account from that of Laplace is the idea that

> The Demon, like a human scientist, must *not* be assumed to ascertain initial conditions with *absolute mathematical precision*; like a human scientist, he will have to be content with a finite degree of precision. [*ibid.*, 34]

As Earman points out, the principle of accountability is imposed to ensure that the required degree of precision for the initial conditions can be known beforehand. [Earman (1986), 8] This is to parry the dismissal of a prediction which fails to meet the specified error limits on the grounds that the initial data were not accurate enough. The conclusions Popper draws from these premisses are somewhat surprising — in two distinct senses. Firstly, Popper reaches the correct (from the point of view provided by hindsight) conclusion that, contrary to widespread belief, classical physics exhibits systems which are not deterministic. This is an important result; it is only that the justification for it seems to be in the wrong track. I will explain this in a moment. Secondly, he infers somewhat impetuously that because the combination of a strong form of instability with the inability of the Demon to ascertain initial condition with mathematically exact precision may lead to breakdown in prediction, determinism must fail also. The conclusion does not follow, if we drop the requirement of prediction and determinism working in tandem. Indeed, in accordance with the ontological predicament of James, some future states of a given system may be unpredictable (given any suitable explication of the Popperian criteria of predictability) although any other future complement of the given process of dynamic evolution than the one fixed from eternity is impossible.

Now, what is wrong with the justification Popper provides for maintaining the thesis that classical physics need not be deterministic? The justification Popper seems to have in mind is the following: classical physics does not imply determinism any more than quantum mechanics does, because there is always an irreducible statistical element in any prediction made with a *prima facie* deterministic theory.[19] So the reason for the breakdown of determinism for Popper is the (practically) unavoidable uncertainty pertaining to statistical analyses of the data provided by observations. It is therefore a purely epistemic matter to assess whether a given phenomenon is under sway of determinism. What complicates Popper's position is that he somehow tends to regard certain theories, e.g., classical physics (the status of which as a theory is not unambiguously decided without certain reservations and conceptual refinements) as intrinsically deterministic (cf. his notion of a *prima facie* deterministic theory above). As Popper's criterion for determinism is epistemic, he encounters severe problems in explicating determinism in the context of classical mechanics. The theory of Newtonian mechanics is interpreted in the Popperian framework by its observational consequences. The semantics of such a theory has to be deductively derivable from the evidence reported (possibly in a observational sublanguage). But as has been shown by Stein, for example, [Stein (1992), (1994)] no such comprehensive theory exists. What is required, is an auxiliary schematization of the observational procedure to help in the derivation of future observations from the theory without requiring the theory itself to account for the entire process of observation. The rather intricate questions of instrumentation and selection of statistical models, among other things, greatly complicate the seemingly straightforward picture of classical mechanics provided by Popper. (Classical physics cannot provide a very realistic picture of the tools and processes of measurement, for example.) The assumption about predictability at any given level of accuracy is too strong to be accounted for, even in the Newtonian picture. Nevertheless, with regard to his conception of *prima facie* deterministic theories, Popper seems to be sliding towards a position resembling Laplace's. As Earman has noted, "Popper's view is particularly awkward in the case of classical statistical mechanics because it has the effect of brushing aside one of the central foundations problems; namely, how can the 'random' and 'chaotic' behavior exhibited on the macro-level by, say, a box of gas be reconciled with the micro-determinism of the gas molecules? After many decades of research, it has become apparent that a large part the answer lies precisely in instability." [Earman (1986), 9] If Popper's aim was, as he claimed in *The Open Universe*, to "make room within physical theory [...] for indeterminism" [Popper (1983), xxi], it has to be admitted that he succeeded — but only according to his own inadequate standards. His success is explainable by his advocating an arbitrary formula of determinism (determinism defined in terms of finite prediction tasks) which was to be couched as a 'scientific' doctrine, scientific determinism, in accordance with his falsifiability criterion. What ultimately challenges this kind of approach — irrespective of

[19]Classical physics is for Popper a *prima facie* deterministic theory on the grounds that its determinism is a property of the theory and not of the world. N.B. It is not at all clear how this conception is to be reconciled with the Popperian account of determinism as predictability. I will touch upon this matter shortly.

being critical of the epistemic point of view — is that in certain well-defined situations, if the Popperian demon is incapable of charting the unfolding of physical events, this could follow, not from the indeterministic nature of the system, but from its instability.[20] Thus, there are varieties of deterministic unfolding of events which Popper's definition is entirely insensitive to. Let me now investigate yet a couple of other definitions which do not seem to have this same defect.

Alternative definitions of determinism

¶In the article "On the Notion of Cause with Applications to the Free-Will Problem" (1953) [Russell (1953)], Bertrand Russell makes the effort of finding a definition of determinism that would not be burdened by the emphatically epistemic elements of Laplace's description. Here's his proposal:

> A system is said to be 'deterministic' when, given certain data, $e_1, e_2, \ldots, e_n$ at times $t_1, t_2, \ldots, t_n$ respectively, concerning this system, if E_t is the state of the system at any time t, there is a functional relation of the form
>
> $$E_t = f(e_1, t_1, e_2, t_2 \ldots, e_n, t_n).$$
>
> The system will be 'deterministic throughout the given period' if t, in the above formula, may be any time within that period . . . If the universe, as a whole, is such a system, determinism is true of the universe; if not, not. [*ibid.*, 398]

Is this a successful, scientifically accurate rendering of the Jamesian vision of determinism? At first glance it would seem so, but when analysed more carefully, it is seen to result in the following position. If we take the example provided by Russell (and reiterated by Earman [Earman (1986), 11]), we can, for the sake of an argument, think of a universe so simple that it contains only a single dimensionless particle. Furthermore, we may suppose that the state of the particle at any instant t is specified by its position coordinates x_t, y_t, z_t. Now, imagine the particle moving in an arbitrary manner, the only condition being that it can occupy only a unique place at a time; the trajectory of the particle may be as complex as you like, provided only that the uniqueness condition is satisfied. Then it is a mathematical fact that there must exist functions f_1, f_2, f_3 such that $x_t = f_1(t), y_t = f_2(t), z_t = f_3(t)$. The upshot of this is that the definition amounts to triviality, no matter in which way you complicate the picture by adding other particles and additional state variables. As Russell noted:

> It follows that, theoretically, the whole state of the material universe at a time t must be capable of being exhibited as a function of t. Hence our universe will be deterministic in the sense defined above. But if this is true, no information is conveyed about the universe in stating that it is deterministic. [*ibid.*, 401]

[20] I will come to the issue of instability in a moment.

Russell tries to evade the triviality result by imposing additional conditions, in the first palace on the function (universally termed *Russell function*), and in the second place on the universe — the latter being a time-honored strategy of genuine meta-physicians. The condition on the Russell function is that it ought to be simple. But how tight is the connection between determinism and simplicity? To a fairly unprejudiced thinker, not that tight. Search for simplicity is a heuristic principle well worth considering in lesser contexts, but when tackling determinism (and the state function of a given system), not a very plausible one. Indeed, one could have either of the following extreme combinations: utmost simplicity of the universe (both in terms of content and temporal evolution) without a trace of the Jamesian variety of determinism on one hand, and a highly complex universe with a completely deterministic pattern of evolution, on the other. How about the latter condition? Russell suggests that time should not be allowed to enter explicitly into the definition of the Russell function. This condition is justified by the independent assumption about the "uniformity of nature", which Russell explains to have the meaning that "no scientific law involves time as an argument, unless, of course, it is given in integrated form, in which case lapse of time, though not absolute time, may appear in the formulae". [*ibid.*, 401] The issue of uniformity, however, is quite complicated. Though it might seem upon the surface that a divergence from uniformity might result in an indeterministic picture of the universe (as in the case of a constant of nature, such as the gravitational constant G, changing linearly as a function of time) it is not at all clear that such dynamics entails non-deterministic evolution that is open to ambiguous future possibilities. It is evident that the problem of time and time symmetries are most intimately connected with the problem of determinism. And they definitely figure prominently within the issue of uniformity. Thus, to give a satisfactory account of the question of "the uniformity of nature", careful scrutiny would have to be given to these issues.[21]

¶Patrick Suppes, whose seminal work in the philosophy of science has proven invaluable in instigating and guiding the study of the issues pertaining to determinism and indeterminism, presents the familiar Laplacian view of determinism in his lectures on 'probabilistic metaphysics' (1974):

> Undoubtedly, the most famous single passage on determinism is that by Laplace, found, I emphasize, not in his *Celestial Mechanics*, but in the introduction to his treatise on probability. He asserts that from a knowledge of the present state of the universe an 'intelligence', or as perhaps we might say in modern terminology, 'a computer adequate to the task', would be able to determine the entire past and future of the universe. [Suppes (1974), 4]

Suppes himself does not subscribe to this view of determinism, of course. He maintains that it is the prevalent view on determinism, and, consequently, one that he should be prepared to criticize in promoting his program of probabilistic empiricism. (This appears on the surface to be the familiar process of making straw-men, although it seems that

[21]It is necessary to note here that it is not the principle of explicating functional relationships as such that is problematic in Russell's account. Rather, the problem verges on the preposterous attempt to encapsulate the whole of physical reality in a single functional relationship (cf. the *Weltformel* of Einstein).

Suppes is here perfectly guileless.) What distinguishes Suppes' approach from that of Popper is that he links the Laplacian version of determinism to a concrete physical *theory*, namely the special system of particle mechanics of which an axiomatic presentation has been given in [McKinsey, Sugar & Suppes (1953)]. The importance of this step cannot be exaggerated. The condition of predictability is not explicitly referred to, although it undoubtedly is tacitly assumed. Most importantly, nothing in Suppes' formulation of determinism hinges on predictability. Instead of a general and indefinite description that Popper gave in connection with his definition of determinism we are now provided with a definite, mathematical theory which purports to exemplify determinism in the Laplacian sense. It is expedient for our purposes to adduce the explication of Suppes. After this we are in a good position to understand the radical transition in our understanding of determinism brought about the three-body problem.

Suppes' system of axioms for the system of particle mechanics is based on five primitive concepts: a finite set P of particles, an interval T of real numbers for the measurement of time, a mass function m, a position function s, and a force function f. The reader who is interested in the exact structure of the system and the content of the axioms is advised to consult the original paper [*ibid.*]. I give here only the theorem that according to Suppes, captures essentially the idea of Laplacian determinism. The theorem is as follows:

Theorem 6 *Let $\mathcal{P} = \langle P, T, m, s, f \rangle$ and $\mathcal{P}' = \langle P', T', m', s', f' \rangle$ be two systems of classical particle mechanics such that $P = P', T = T', m = m'$ and $f = f'$, and for some t_0 in T let the positions and velocities of identical particles in the two systems be identical. Then the trajectories of the particles in both systems are identical or, in other words, the position functions are identical for the two systems, i.e. $s = s'$. [Suppes (1974), 6]*

Note that the theorem depends on the absence of collisions. Thus it is necessary to make either of the following assumptions for the theorem to hold: (1) particles are penetrable,[22] (2) during the period of time T for which the analysis holds, no collisions occur, either because of separation of distance or because of the particular configuration of motion. This theorem is, in essence, equivalent to the definition of Laplacian determinism which John Earman gives in his book [Earman (1986)]. The only difference is that Earman utilizes the "pictorially appealing approach" of possible worlds. In Earman's nomenclature $\mathcal{W}$ stands for the collection of physically possible worlds (possible worlds satisfying the natural laws obtaining in the actual world). The definition runs:

Definition 7 *The world $W \in \mathcal{W}$ is* Laplacian deterministic *just in case for any $W' \in \mathcal{W}$, if W and W' agree at any time, then they agree for all times. [*ibid.*, 13]*

[22] Although not a tenable assumption within classical physics, this is not as unreasonable a restriction as it sounds: for example, two photons cannot couple together directly, since they carry no charge. This means, roughly, that in a certain sense a system of photons can be regarded as penetrable particles. However, there are higher order processes in which photon-photon interactions play a very important role, and hence the assumption of unobstructed penetration is not universally valid even in the case of photons. Neutrino-neutrino interactions are even more exotic (and do not manifest at all within the ordinary energy range encountered in the solar system, for example), taking place mainly in core-collapse supernovæ, Early Universe, and possibly in gamma-ray bursts. These are the best examples I can think of in connection with the assumption of free penetration.

Other varieties of determinism could be produced indefinitely by explicating the features, and modifications of those features, of the space-time regions which are determined, and which, in their turn, do the determining. These varieties would turn out useful in discussing the status of determinism in relativistic particle mechanics [*ibid.*, pp. 17 & 55–79], but since my focus in this section is on classical celestial mechanics, I do not examine their characteristics here, but turn to matters celestial instead.

The three-body problem — its historical background

The notorious three-body problem, "the most celebrated of all dynamical problems" [Whittaker (1937)], can be presented deceptively simply: How do three celestial bodies move under their mutual gravitational attraction? To be a little more specific, this may be enunciated as follows (as is done by one of the most prominent mathematicians of the beginning of the last century):

> Three particles attract each other according to the Newtonian law, so that between each pair of particles there is an attractive force which is proportional to the product of the masses of the particles and the inverse square of their distance apart: they are free to move in space, and are initially supposed to be moving in any given manner; to determine their subsequent motion. [*ibid.*, 339]

The simplicity of the statement of the problem — as is often the case in mathematics — belies the complexity of the issues involved. Although the corresponding problems of one and two bodies are exactly solvable in the sense that there is a solution in closed form in terms of elementary functions for both of them, the problem of three bodies is a difficult, nonlinear problem for which no similar type of solution exists. In addition to being a fascinating, well-posed problem of pure mathematics, the three-body problem has had considerable interest for potential solvers because of its relevance for the question about the stability of the solar system. According to June Barrow-Green, between 1750 and the beginning of the 20th century, more than 800 papers relating to the problem were published. [Barrow-Green (1997)] The first particular solutions to the problem were found by Lagrange in his prize memoir, *Essai sur le Problème des Trois Corps*, which was submitted to the Paris Academy in 1772. The solution Lagrange found is universally called the Lagrangian equilateral triangle solution. Lagrange's method consisted of dividing the problem into two parts: (a) the determination of the mutual distances of the bodies, and (b) having solved (a), the determination of the plane of the triangle in space and the orientation of the triangle in the plane. Furthermore, he proved that if the part (a) could be solved, the part (b) could be solved also. Other prominent characters who worked on the restricted three-body problem included Jacobi (1836)[23], Tisserand (1889, 1896) (who studied the problem in connection with devising methods for identifying comets), Hill

[23] In parentheses is the year of the original publication of the result provided by the particular character in question.

(1878) and Delanauy (1860). The so called 'classical' period of the studies on the problem reached its final phase with Poincaré (1892-1899).[24]

In spite of the successes in tackling the special cases of the problem, the solution of the general three-body problem remained elusive even after two-hundred years following the publication of Newton's *Principia*. In the general problem, as adduced above, all the masses of the three bodies are non-zero and their initial positions and velocities are arbitrary, i.e. they might not be arranged in any particularly symmetric way. This implies that there are not any coordinate transformations that would significantly simplify the problem. A large part of the mathematical difficulties derives from this fact. In the two-body problem, for example, the solutions are most easily found by utilizing the center of mass coordinate system. In the two-body problem and in the restricted three-body problem the situation can be dynamically reduced to a consideration of inverse square force fields with the lines of force passing through the center of mass of the system. In the general case this is not possible: the lines of mutual forces do not pass through the center of mass of the system. This means that the motion of each body has to be considered in conjunction with the motions of the two other bodies. Before the age of powerful computers, this was the primary cause for the problem's intractability.

The problem was considered so important and difficult that at the suggestion of leading scientists, the King of Sweden and Norway, Oscar II established a prize for its solution to mark his 60th birthday on January 21, 1889. The condition for the solution was that it should be in the form of a series expansion giving the positions of the three bodies at all future moments of time following an arbitrary starting configuration. The competition achieved unusual recognition that stretched well beyond the circles of mathematics. This has undoubtedly been due to a combination of royal patronage and carefully planned public relations, as Barrow-Green has pointed out. [Barrow-Green (1997), 49] But the personality of King Oscar II may have been a considerable factor also, as the following excerpt from Sonya Kovalevskaya's autobiography illustrates:

> King Oscar is a pleasant and cultivated person. As a young man he attended lectures at the university, and still today shows an interest in science, although I cannot vouch for the profundity of his erudition. He has no official contact with the university but is extremely sympathetic to it and very amicably disposed towards its professors in general and to myself in particular. [Kovalevskaya (1978), 228]

Taking all these factors together, the publicity gained by the competition was deemed to result in the universal celebration of the mathematician who would solve the problem. For many years the prize remained unclaimed. The prize was eventually awarded to

[24] In the restricted three-body problem the mass of the third particle is assumed to be 0. In such a configuration, the energy function of the massless particle is $H(x, y, z, \dot{x}, \dot{y}, \dot{z}) = \frac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) - \omega(x, y, z)$, where the last term is the effective potential $\omega(x, y, z) = \frac{1}{2}(x^2 + y^2) + \frac{1-\mu}{r_1} + \frac{\mu}{r_2}$. The so-called *Hill's region* is defined as the projection of $H(x, y, z, \dot{x}, \dot{y}, \dot{z}) =$ constant into position space (x, y, z) and its surface is called a zero velocity surface. The topology of the Hill's region depends on the energy level.

Poincaré in 1899 on the basis that he was thought to have made most progress in the subject. What strikes one as particularly ironic in this is that Poincaré was the one least convinced of the suitability of the series expansion method in solving the problem. In fact, he severely criticized such methods. Poincaré did not solve the given problem. It was only after twenty odd years that Karl F. Sundman completed the given task. However, the series expansion given by Sundman proved to be useless for the purpose of calculating orbits: his expansion converged so slowly that it was of no practical help in doing the calculations. Nevertheless, a more fundamental objection to Sundmans' method became evident only much later. It turned out that Poincaré had been on the right track from the start: as modern computer calculations of orbits have verified, the dynamical laws governing the orbits bring about chaotic behavior and thus any deterministic series expansions thought to describe the motion are wide off the track.

Poincaré's results pertaining to the problem of three bodies were published in two separate memoirs, the first of which was submitted for the Oscar's competition, and the second of which was a revised version of the competition memoir.[25] The paper of Poincaré which became widely applauded was not, curiously enough, the one which won the Oscar competition, but the one that was a revised version of the original. Poincaré made substantial corrections and alterations to the original but gave no indication of the extent of his alterations. This was to arouse some confusion in the mathematical community for many years, since it was widely believed that the revised version was in fact the one that won the competition. Only the discovery of a printed version of Poincaré's original memoir has made it possible to reconstruct the nature of his alterations. [*ibid.*, 1] The issues Poincaré tackled with in his memoirs can be briefly depicted as follows. The most important concept Poincaré focused on (although he discussed several mathematical concepts within the memoirs) is that of Hamiltonian systems with two degrees of freedom. This concept is intimately related to the restricted three-body problem. In developing the theory, a major role is taken up by the topics of invariant integrals and periodic solutions. It is in these memoirs that Poincaré for the first time takes the issue of invariant integrals under careful scrutiny and thereby gives a detailed account of them along with his famous recurrence theorem. Furthermore, the notion of periodic solutions is for the first time shown to be a powerful tool for studying the qualitative questions pertaining to the theory of (ordinary) differential equations. In connection with this issue, the most important role is seen to be bestowed on his theory of asymptotic solutions. All of these theoretical questions are tackled in the first parts of the two memoirs. The application of the theory of asymptotic solutions to the restricted three-body problem is taken up in the second part of the memoirs. Poincaré proceeded by a series of approximations. This implied that Poincaré was enforced to take into account an increasing number of terms in the power series expansion for the solution and then interpret the geometrical nature and significance of these improved approximations. The geometrical point of view was important, because it is the geometrical description of the asymptotic solutions that is most dramatically affected by the discovery and correction of the error in the series ex-

[25] It is not possible for me to give a detailed elaboration of Poincaré's work on the three-body problem here. June Barrow-Green's book [*ibid.*] provides an illuminating historical and topical account of the problem, and anyone who is interested in the details of Poincaré's work should consult this book.

pansion. Nevertheless, the essential changes that the correction entails are in fact due to a conjunction of errors arising in two separate parts of the preceding analysis. What was the upshot of these seminal mathematical insights? Barrow- Green puts this most succinctly:

> Briefly, in his original account Poincaré did not draw a distinction between autonomous and nonautonomous Hamiltonian systems of differential equations. As a result he drew mistaken conclusions about the convergence of the series used to describe the asymptotic solutions of the problem. Originally he had believed that the series were convergent and led to asymptotic trajectories with behavior which he could easily understand. In the revised memoir he showed that the series were actually asymptotic expansions, and, with his discovery of homoclinic points, he found that the behavior of the trajectories was anything but easy to describe. In fact the behavior was what today would be called chaotic. Thus, contrary to what is sometimes thought, Poincaré did not win the Oscar prize for his discovery and analysis of the behavior what he called doubly asymptotic solutions (and later called homoclinic solutions), that is, the 'chaotic' trajectories, but rather for the underlying theory which eventually led to his correct description of these solutions. [*ibid.*, 3]

It is well known that Poincaré continued his work in celestial mechanics after the publication of the first two memoirs, and the period of prodigious productivity was finally crowned in the famous multi-volume treatise *Méthodes Nouvelles* published between 1892-1899. After that, the road was open for the most fascinating developments in mathematics but only in principle. It took many years before mathematicians started to appreciate the real significance of Poincaré's results, but once the depth of his findings was realized, it resulted in an exultant period of unfathomable creativity. The appreciation of Poincaré's work has grown steadily thereafter. Quite recently Philip Holmes, whose book (co-authored with John Guckenheimer) on dynamical systems is a landmark in the field, has described Poincaré's memoir as: "... the first textbook in the qualitative theory of dynamical systems." [Holmes (1990)] I will now turn to study the outlines of the development in another branch of mathematics which ultimately was brought to bear heavily on the issues related to classical mechanics, namely, randomness.

What is randomness?

Randomness is an elusive concept. A first approach to enquire its meaning on an everyday level brings various images of different stochastic phenomena, such as tossing of a coin, to mind. The more philosophical question to ask is: what is common to these different stochastic processes? Is there possibly a criterion of determinacy for the concept of "randomness"? Philosophers have traditionally thought the concept too vague to be captured by any adequate definition. A notable exception is Richard von Mises (the Viennese mathematician and philosopher), one of the first to suggest a way to approach the definability problem. Following the pioneering work of von Mises, independent attempts to define randomness have been made among others by A.N. Kolmogorov and Per Martin-Löf. [Martin-Löf (1966), (1970a), (1970b), 1971][26] The fundamental idea lying behind the efforts of von Mises, Kolmogorov and Martin-Löf to define randomness is, roughly, to make explicit the criteria for determining whether a given sequence (of numbers) exemplifies randomness. By devising tools to study the properties of a given sequence, each of them arrives at a particular notion of a *random sequence*. What does this idea of a random sequence consist in? Inspired by Richard von Mises's work relating to randomness A.N. Kolmogorov devoted his later life to investigating the problem of the foundations of probability. Concerning the meanings conferred to the term "randomness" he once remarked:

> In everyday language we call random these phenomena where we cannot find a regularity allowing us to predict precisely their results. Generally speaking, there is no ground to believe that a random phenomenon should possess any definite probability. Therefore we should have distinguished between randomness proper (as absence of any regularity) and stochastic randomness (which is the subject of probability theory). There emerges a problem of finding the reasons for the applicability of the mathematical theory of probability to the real world. [Kolmogorov (1984)]

Kolmogorov is here making a very important distinction between two kinds of randomness. I alluded above to the very popular idea of associating randomness with a stochastic process. But this is seen to be too narrow a view; according to Kolmogorov, irregular sequences are distinguishable from those which show irregularities and statistical regularities (as the stochastic processes do) by the following property: in the latter type of sequences, the *Kolmogorov-complexity* of an initial segment divided by the length of that segment tends to stabilize. This amounts to the kind of behaviour that can be described by the usual notions of statistics and probability, such as the normal distribution and standard deviation, for example.

I briefly explain the meanings of the relevant notions. Kolmogorov– complexity is usually defined as follows:

Definition 8 *Let $A : 2^{<\omega} \to 2^{<\omega}$ be a partial recursive function with Gödel number $\ulcorner A \urcorner$. The complexity $K_A(w)$ of w with respect to A is defined to be*

$$K_A(w) = \begin{cases} \infty & \text{if there is no } p \text{ such that} \quad A(p) = w \\ \mid p \mid & \text{if } p \text{ is a shortest input such that} \quad A(p) = w. \end{cases}$$

Here A means effectively a rule which produces a given sequence from its code p. The intuitive idea is that the complexity of a word w with respect to a rule A is the *length of*

[26] Their approaches are masterfully surveyed in the doctoral dissertation of van Lambalgen. [van Lambalgen (1987)]

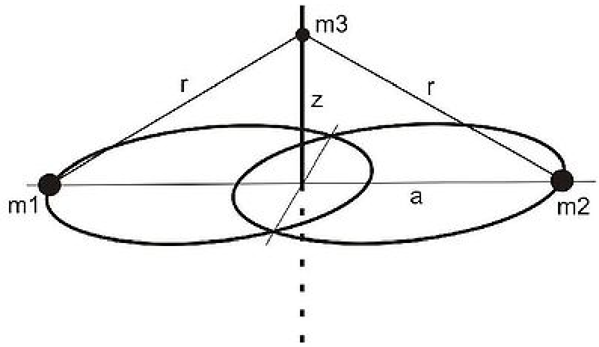


Figure 6.1: Sitnikov Problem Configuration.©Wikipedia.

the shortest input (code) p such that $A(p) = w$. This is a very elegant and powerful concept, indeed. In a nutshell, then, it can be stipulated that a sequence is random if it is of maximal complexity. Intuitively this means that a genuinely irregular sequence cannot be compressed to a representation the length of which is less than that of the original sequence. Kolmogorov's work on complexity and randomness is far too sophisticated to be adduced here in its full richness. Nevertheless, I feel that the rough depiction of randomness I have given above has prepared us to move on to study an interesting example of the three-body problem, where the questions of determinism, randomness and chaos all come together.

Randomness in a deterministic system — the three-body problem

In what follows, I will present a special case of the three-body problem which very clearly displays the intrinsic chaotic features characteristic of instable dynamical systems. The intuitive idea of instability is this: wide divergence in the behavior of two systems identical except for initial conditions is observed even when the initial conditions are extremely close. There are two sources of difficulty as regards the prediction of the behavior of unstable systems. The first difficulty results from the fact that the initial conditions of a given system can be measured only approximately. This leads to the impossibility of predicting the behavior of the system for any but short intervals of time. Predictive failures are then attributed to possible inaccuracies in the initial conditions. This difficulty I have already touched upon in connection with Laplacian determinism. The second source of difficulty is mathematical. It might happen that the solutions are not representable in a closed, analytical form. A usual procedure, then, is to use various approximative techniques, of which the series expansion methods are the most common. The problem with the series expansions is that they do not allow for accurate predictions. The accumulation of small errors in numerical methods of approximations can lead to unavoidable problems of accuracy. These both aspects are relevant for the following example.

The special case of the three-body problem I would like to consider is this:[27] There are two particles of equal mass $m = m_1 = m_2$ moving according to Newton's inverse-square law of gravitation in an elliptic orbit relative to their common center of mass which is at rest. The third particle has a nearly negligible mass ($\frac{m_3}{m} \to 0$), so it does not affect the motion of the other two particles, but they affect its motion. This third particle is moving along a line perpendicular to the plane of motion of the first two particles and intersecting the plane at the center of their mass — let this be the z-axis. From symmetry considerations we can see that the third particle will not move off the line. The restricted problem is to describe the motion of the third particle. To obtain a differential equation in simple form, we normalize the unit of time so that the temporal period of rotation of the two masses in the x, y-plane is 2π, we take the unit of length to be such that the gravitational constant is one, and finally $m_1 = m_2 = \frac{1}{2}$, so that $m_1 + m_2 = 1$. The force on particle m_3, the particle of interest, exerted by the particle 1 is:

$$\mathbf{F}_1 = -\frac{m_1}{z^2 + r^2} \cdot \frac{\mathbf{r}_{31}}{\sqrt{z^2 + r^2}}, \tag{6.1}$$

where r is the distance in the x, y-plane of particle 1 from the center of mass of the two-particle system m_1 and m_2, and this center is, of course, just the point $z = 0$ in the x, y-plane. Note that $\frac{\mathbf{r}_{31}}{\sqrt{z^2+r^2}}$ is the unit vector of direction of the force $\mathbf{F}_1$. Similarly,

$$\mathbf{F}_2 = -\frac{m_2}{z^2 + r^2} \cdot \frac{\mathbf{r}_{32}}{\sqrt{z^2 + r^2}}, \tag{6.2}$$

So, simplifying, we obtain as the ordinary differential equation of the third particle:

$$\frac{d^2 z}{dt^2} = -\frac{z}{(z^2 + r^2)^{\frac{3}{2}}}. \tag{6.3}$$

There are fundamental analytical difficulties entwined with this easily describable problem, but some basic results are quite understandable in informal terms. One special case deserves particularly to be mentioned. Near the escape velocity for the third particle (the velocity at which it leaves the system and does not periodically return) the periodic motion is very irregular. There is a very remarkable theorem that can be proved about this periodic motion. If we denote by $t_1, t_2, \ldots$ the times at which the particle intersects the plane of motion of the other two particles, then the following theorem about the largest integer s_k equal to or less than the difference between t_{k+1} and t_k multiplied by a constant[28] can be proved:

Theorem 7 *Given that the eccentricity of the elliptic orbits is positive but not too large, there exists an integer, say α, such that any infinite sequence of terms s_k with $s_k \geq \alpha$, corresponds to a solution of the deterministic differential equation governing the motion of the third particle.*[29]

[27] The presentation of the problem is adapted from Patrick Suppes. [Suppes (1987), 243–245]

[28] This constant is the reciprocal of the period T of the motion of the two particles in the plane.

[29] This theorem is due to the Russian mathematicians Sitnikov and Alekseev.

In the framework of the theory of random sequences, the following corollary immediately follows:

Corollary 1 *Any random sequence with terms $s_k \geq \alpha$ corresponds to a solution of the deterministic differential equation governing the motion of the third particle.*

Especially striking is the possibility of interpreting this sequence as the random sequence of heads and tails. We only have to pick two integers greater than α to represent this sequence — one representing heads and the other tails. This illustrates in a most dramatic way that the motion of the particle is completely unpredictable even though deterministic. For each random sequence there exists a set of initial conditions that determines the corresponding solution. What is worth pointing out is that, contrary to the actual experiment of flipping a coin, there is no distribution over initial conditions and thus no uncertainty about them in the problem here considered. And conversely, no single trajectory in the coin-flipping case exhibits in itself such random behavior.

Summary

Now, what are the implications of these observations for evaluating the status of determinism? I think the above example points clearly towards a now familiar answer: the fact that the same phenomena can be both deterministic and random, forces us to revise our presuppositions about the importance of the classical determinism/indeterminism distinction. It seems evident on the basis of examples like the one adduced above, that randomness is, in one sense, just a feature of the most complex deterministic systems. Indeed, the most sophisticated definitions of randomness are given in terms of complexity, as I noted above in connection with randomness. Let me reiterate the idea of complexity in somewhat different terms: the most primitive sense of complexity reduces to the classical computational definition of complexity. According to this definition the complexity of a sequence of integers is associated with the minimal computational capabilities necessary to reproduce it. More exactly, the complexity of a sequence of infinite symbols is measured by the length of a minimal computer program that will generate the sequence. According to this definition, random sequences are of maximal complexity. This definition is effectively equivalent to the definition of Kolmogorov-complexity given above. The particularly important implication of these observations is the separation of determinism and predictability. The most complex deterministic systems are totally unpredictable in their behavior. We are then left in a situation where we cannot distinguish between determinism and indeterminism (as traditionally conceived). This is one of the lessons to be learned from the complex behavior of the classical dynamical systems.

This particular problem in the interdisciplinary cross-section of mathematics, computer-science, physics and philosophy is a fine example of the kind of concept-forming processes that fall under Carnap's general program of explication. It is precisely genealogies of this kind (the above presentation providing only a sketch of such a genealogy), that are needed to back up the more detailed tasks of explication that have recourse to specific methods of formalization. In this particular example of a survey of the relations between determinism, indeterminism and randomness we have all the characteristics of a process of *clarification* that is an essential prerequisite of Carnapian explication. In fact, we have (1) a sufficiently well-defined problem stated in terms of a relatively precise language that we can reach mutual agreement on, (2) various, well-founded strategic options to tackle the problem in question, (3) a variety of languages (formal systems, professional idiolects, etc.) in which to state the strategic options exactly, (4) a dialectic among the different suggestions, and (5) a convergence towards a commonly accepted *exact* formulation of the problem and specific techniques for its solution that are taken as the standard *for the time being*. Of course, many other problems in science are tackled in a similar manner. It is only that the problem of randomness seems to provide a particularly illuminating example of the process of clarification and – more importantly – one of continuing interest among contemporary scientists.

The domain of philosophical history of science is, of course, vast. To orient oneself in the menagerie of different histories provided of particular *explications*, and to acquire the skills needed to devise such histories, one would certainly benefit from some general guidelines as to the general methodology of history. The level of methodological self-consciousness of philosophers in this vein has not been particularly high, especially not in the area of history of natural sciences. This is one reason for attempting to make more explicit the tools that are available for philosophers (and historians) who take up the important task of Carnapian clarification. Such tools may then serve the purpose of (descriptive) pragmatics which we have seen as a *desideratum* of modern philosophy of science.

Chapter 7

CONCLUSION

"Ε᾿ις ὕδορ πάντα ἀνάλυεσθαι." *"everything dissolves in water"*
– Thales

7.1 Carnap as an Exemplar of the Scholar's Vocation

In the preceding pages I have provided a panorama of the range of problems and questions that constitute the living kernel of Rudolf Carnap's philosophy. These comprise an honorable collection including the nature of *pseudo-problems* in philosophy, the foundations of mathematics, physicalism and the unity of science, the logical syntax of language, the nature of empiricism, (pure) semantics, language planning, probability and inductive logic, the status of theoretical language, and problems concerning values and practical decisions. Rather than focusing on a particular class of problems, I have attempted to discern an overarching thematic and a certain continuity in Carnap's thought. One relatively stable feature of Carnap's philosophy is the reliance on particular methodological tools. The body of this work concentrates on the development and applications of the method of analysis in Carnap's philosophical work during the years 1922-1950. The apogee of this development is the method of explication, the most elaborate and systematic characterization of which is found in the first chapter of Carnap's *Logical Foundations of Probability* [Carnap (1950b)].

Another stable feature of Carnap's philosophy is covert; the overall motivation behind his philosophy. In keeping with the cosmopolitan cultural atmosphere of the Youth Movement in Germany and the later radical utopianism of the Vienna Circle, Carnap remained a political radical throughout his life. It was mainly due to adverse circumstances that he never clearly spelled out his political motifs. In his Autobiography he ventured to provide the following account of his moral and political convictions:

> The view that recognition of the non-cognitive nature of value-statements is either conducive to or symptomatic of a loss of interest in moral or political

problems seems clearly refuted by my own experience. I have maintained the thesis for about thirty years. But throughout my life, from my childhood to the present day, I have always had an intense interest in moral problems, both those concerning the life of individuals, and, since First World War, those of politics, but I was always interested in political principles and I have never shied away from professing my point of view. All of us in the Vienna Circle took a strong interest in the political events in our country, in Europe, and in the world. These problems were discussed privately, not in the Circle which was devoted to theoretical questions. I think that nearly all of us shared the following three views as a matter of course which hardly needed any discussion. The first is the view that man has no supernatural protectors or enemies and that therefore whatever can be done to improve life is the task of man himself. Second, we had the conviction that mankind is able to change the conditions of life in such a way that many of the sufferings of today may be avoided and that the external and the internal situation of life for the individual, for the community, and finally for humanity will be essentially improved. The third is the view that all deliberate action presupposes knowledge of the world, that the scientific method is the best method of acquiring knowledge and that therefore science must be regarded as one of the most valuable instruments for the improvement of life. In Vienna we had no names for these views; if we look for a brief designation in American terminology for the combination of these three convictions, the best would seem to be "scientific humanism". [Carnap (1963), 82–83]

Relating this discussion to the more explicitly scientific aspects of Carnap's philosophy, it is easy to see that Carnap's method of explication potentially provides a framework for such a politically motivated improvement of the condition of humanity. Carnap never worked out the broader implication of his method of explication, the range of problems that it would enable one to address, and the possibilities thus opened for systematic social planning. *This* is the motivation behind André Carus' interpretation of Carnap's ideal, i.e., to attempt a *reconstruction* of the Carnapian method of explication as a vehicle of Enlightenment. I have critically examined the suggestion of Carus within an interpretive framework that adds a novel historical dimension to the discussion. I argue that the Principle of Tolerance which played such a prominent role in modifying and catalyzing Carnap's thought about the method of explication, merits careful scrutiny, both as a philosophical principle (a task somewhat neglected to this date) and as one of the constitutive notions of the program of Enlightenment. Although widely acknowledged as an important regulative principle in Carnap's system, Carus' interpretation, to my knowledge, is the only one that ventures to press the implications of Carnap's basic idea to their limits. Even so, Carus' interpretation leaves one crucial aspect of the Principle unexamined, viz., its *moral* underpinnings. I have argued that a full recognition of the importance of Carnap's Principle of Tolerance requires a careful scrutiny of the value-theoretical background in which it is embedded. Carus eschews this task. Even though there are some definite aspects to Carnap's ideal that render it similar (in spirit but not in letter) to the one advocated by the harbingers of Radical Enlightenment, the interpretation of the ideal of explication as a modern vehicle of Enlightenment seems to me impetuous. The fundamental characteristic of the classical Radical Enlightenment systems of thought, most notably the one of Spinoza, is the harmonious confluence of intellectual and moral dimensions of Reason. Such a confluence must, in order to be effective, be backed up with an ethical *theory*.

The remaining question is, then, whether Carnap in fact had such a value-theoretical background to draw on. This is bound to remain an open question as long as no explicit documentary evidence about Carnap's explicit ethical views can be found in Carnap's *Nachlass*. Nevertheless, the general idea about explication as an instrument of improvement of the human condition is very attractive. It is also surprisingly close to the view of R. G. Collingwood who envisioned a gradual betterment of human knowledge through a process of evolution which he called "a scale of forms". Collingwood explicitly identified these scales with series of terms in dialectical relationship. A key feature of a dialectical relationship is the distinction between implicit and explicit, where the the consecutive terms in a series are related so that the later terms make explicit what remains only implicit in the earlier terms. Collingwood reports of this insight dawning on him in *Speculum Mentis*:

> I may perhaps be permitted here to refer to a book called *Religion and Philosophy* which I published in 1916, and in which I tried to give a general account of the nature of the religious consciousness, tested and illustrated by detailed analyses of the central doctrines of Christianity. With much of what that book contains I am still in agreement; but there are certain principles which I then overlooked or denied, in the light of which many of its faults may be corrected. The chief of these principles is *the distinction between implicit and explicit*. I contended throughout that religion, theology and philosophy were identical, and this I should now not so much withdraw as qualify by pointing out that the 'empirical' distinction (i.e. real but unexplained) between them is that theology makes explicit what in religion as such is always implicit, and so with philosophy and theology. This error led me into too intellectualistic or abstract attitude towards religion, of which many critics rightly accused me.[1]

Although the framework in which Collingwood presents his thesis – concentrating explicitly only on the relations between religion, theology and philosophy – is quite foreign to the Carnapian framework of explication, there is a family resemblance. Collingwood intended his theory of scale of forms to be applicable in all compartments of human knowledge. The essential insight of Collingwood is, then, that the effort of the conscious mind to attain clarity and distinctness ruminates in every conceivable area of human activity, including the delicate problems pertaining to values and practical decision making. Similarly, Carnap's method of explication provides the broad framework of addressing such questions, given the qualifications about the value-theoretical background of his method.

[1][SM 108n] My italics.

The tension between the characteristic *habitudes* of both Enlightenment and Romanticism which manifests itself quite clearly in Carnap's intellectual development raises once more the question whether any thinker who has interests as wide as Carnap, can straightforwardly be classified as a representative of a singular intellectual tradition. The contemporary trend in Carnap scholarship has very much been along the lines of an attempt to focus on those aspects of his thinking that embrace open-mindedness, tolerance and pragmatism in contrast to the (already *passé*) varieties of interpretation which forged Carnap a solid reputation as a philosophical dogmatist in the mid-twentieth century. More recently, the general contention has been that the more one engages with Carnap's 'real' thought, the more clearly one discovers elements of philosophical pluralism and tolerance for intellectual diversity in general. Indeed, the revisionist picture of Carnap as the unyielding bridge-builder who is always looking for possibilities to reconcile seemingly irreconcilable philosophical positions reaches its culmination in the bold conjecture of André Carus who reads Carnap as essentially promoting the ideal of Enlightenment in our modern society. This ideal, quite true, is not any of the original forms of Enlightenment ideals that flourished in the Western Europe in the seventeenth and eighteenth centuries. These are all obsolete by now and their rehabilitation *per se* would seem quite absurd. Carus explicitly states this as he pleads for a modern, Carnapian variety of Enlightenment:

> The Enlightenment has traditionally claimed that there are *impersonal procedures* that can effectively determine what makes a given piece of knowledge 'better' (in a more or less precisely specifiable way) than another. [...] And indeed, it has proven much more difficult than anyone had ever imagined to specify any such procedures (as the failure of Carnap's own lifelong efforts eloquently testifies). The Enlightenment's heavy reliance on them is clearly no longer tenable in its original form. The original Enlightenment, though ethically and politically pluralistic, was *intellectually* not pluralistic. It really did think that natural science — including social science, in its view — would, progressively, find the true answers. Science, it thought, would provide the replacement, once and for all, for the lore of the church and other traditional institutions, thereby providing *authoritative* knowledge, in an authoritative (encyclopaedic) language, in the ideal future. [Carus (2007), 307–308]

This ideal can be justifiably be said to have been discredited. But Carus assures his readers that the "Carnapian ideal can correct this shortcoming by building in a radical pluralism at the very heart of the Enlightenment project." [*ibid.*] This is not at all as evident as Carus wants to make us believe.

7.2 The Tension Between the *Inner* and the *Outer* in Carnap's Thought

As we have seen, there is an essential tension between the diametrically opposite currents of Enlightenment and Romanticism which both figure prominently in Carnap's development. Firstly, the prospects of Enlightenment thought in the post-world-war Germany were quite bleak and unrelenting, as the many varieties of anti-intellectualism and irrationalism of that period attest. We have already referred to these phenomena and the influence of *Lebensphilosophie* upon German intellectual culture at large. Philosophically there was, however, a more systematic reason for the disdain that was felt towards Enlightenment thought. For example, John Dewey had already remarked in *German Philosophy and Politics* [Dewey (1915)] on the difficulties upon which German philosophy had stumbled because of the "systematic intellectual error" the essence of which he diagnosed as being too close an adherence to a *dualistic* interpretation of Kant's philosophical architectonics, according to which there are two strictly separated realms, "one outer, physical and necessary, the other inner, ideal and free". [*ibid.*, 28]. The dualism between *Geist* and *Leben* which was a hallmark of the more superficial strands of German philosophical thought had its impact also on Carnap's philosophical development. The tension between the *inner* and the *outer* which he was never able to release satisfactorily, drove Carnap on philosophical paths that would drift further and further apart. Outside observers, like Dewey, saw the situation in German philosophy as symptomatic of a more general cultural influence:

> since Kant's times [it] set its intellectual and spiritual clocks by the Kantian standard: the separation of the inner and the outer, with its lesson of freedom and idealism in one realm, and of mechanism, efficiency and organization in the other. [...] It does seem true that [...] Germans [...] can withdraw themselves from the exigencies and contingencies of life into a region of *Innerlichkeit* which at least seems boundless. [Dewey (1915), 45]

Celebration of the inner which manifested most clearly in neo-romantic art, music and poetry was possible also in a purely theoretical realm. This was fully evident to Carnap who saw the "the ocean of unlimited possibilities" of the newly discovered beautiful formal systems in logic and mathematics as an independent, pure and solid region of *Innerlichkeit*. This element of German philosophical culture in Carnap's thought has been most succinctly described by Thomas Mormann as "formal romanticism". The purely formal interpretation of the realm of the possible was, however, marked by a certain intellectual stubbornness, which was to characterize Carnap throughout his life. As Mormann puts it:

> Complementarily to his predilection for exploring formal possibilities, throughout his intellectual career Carnap had no sense for the "messiness" of the practical realm. Scientific matters that pointed in this direction he delegated

> to disciplines such as psychology, sociology, or history — he himself was always really interested only in the pure realm of philosophy of science as logic of science. He never showed any sympathy for matters of approximation, vagueness, and ambiguity and never took seriously Neurath's pet idea that "Ballungen" were inevitable even in our best science. [Mormann (2010), 11]

In Mormann's view, Carnap's predilection for purely formal and theoretical questions in the theory of science, his relentless quest for a true *Wissenschaftslogik*, was more akin in spirit to the Romantic influences operating in the Wilhelminian pre-world-war Germany than the Enlightenment ideals perpetuated by the French *philosophes* or their successors, like Condorcet, Saint-Simon or Auguste Comte, who all exemplified the engineering and revolutionary spirit of the early ninenteenth-century *École polytechnique*. Mormann gives Nietzsche a pride of place among the romantic philosophers who had a most powerful impact on intellectual culture in Germany. It is true that Carnap's generation was heavily influenced by Nietzsche, but to boldly claim that even Carnap himself was deeply influenced by Nietzsche's powerful and visionary writings, seems *prima facie* impetuous. Gottfried Gabriel, who has classified philosophical styles along a continuum which has the fixed points of poetry and science as its two diametrically opposed poles, described the tension in Carnap's philosophy in the following terms: "For Carnap, Frege's *Begriffsschrift* lied on the desk, so to speak, and Nietzsche's *Zarathustra* on the bedside table" (cf. [Gabriel (2004), 12]). Mormann wants to go beyond this, asserting that "Nietzsche was more than just a metaphysical poet (*Begriffsdichter*) who expressed the *Lebensgefühl* of Carnap's generation in unequaled rhetorical elegance and intensity. Nietzsche influenced considerably his thought-style and even the content of philosophizing." [*ibid.*, 12] Unfortunately, Mormann does not provide in his paper but only a couple of examples to back up his claim. Admittedly, a thorough examination of the philosophical relationship between Carnap and Nietzsche would require a far more lengthier treatment than is possible in a paper. Nevertheless, Mormann has opened up an interesting topic for academic discussion and his interpretation of the Carnap–Nietzsche connection is likely to raise a lengthy debate among Carnap scholars. At the very least Mormann has shown that the dialectic between the Enlightenment and Romanticism in Carnap's thought is much more delicate than has hitherto been admitted.[2] Remember the remarks Nietzsche made on the thematic of objectivity and its many varieties, both in the arts and sciences: "There is required above all great artistic facility, creative vision, loving absorption in the

[2]The salient examples Mormann provides in this connection are (i) the manuscript *Vom Chaos zur Welt* [Carnap (1921/22)] which Carnap himself considered as the "nucleus of the *Aufbau*", (ii) "Überwindung der Metaphysik durch logische Analyse der Sprache" [Carnap (1932)] and (iii) "Theoretische Fragen und praktische Entscheidungen" [Carnap (1934)], all of which are claimed to have been influenced by Nietzsche. Mormann states that "[i]n *Chaos* Carnap subscribed to a pseudo-Nietzschean 'will to order' (for him apparently more appealing than the original 'will to power) that was the 'irrational starting point' of the orderly constitution of the world which the philosopher attempted to realize. In Aufbau, Carnap quoted several times approvingly a rather apocryphal edition of Nietzsche's *The Will to Power* (edited by Max Brahn) (*Aufbau* §§65, 67, 163))." Furthermore, "Overcoming Metaphysics by Logical Analysis of Language" rehearses, according to Mormann, a key theme of *Zarathustra*, to wit, "overcoming" and "self-overcoming". And finally, "[i]n 'Theoretische Fragen und praktische Entscheidungen' (Carnap 1934) Carnap raged against theology and metaphysical philosophy as 'dangerous narcotics having a detrimental effect on reason' in a way that reminds one not only of Marx but also on Nietzsche (cf. *The Gay Science*, Book 3, 147)." [*ibid.*]

empirical data, the capacity to imagine the further development of a given type — in any event objectivity is required, but as a positive quality. So often objectivity is only a phrase. Instead of the outwardly tranquil but inwardly flashing eye of the artist there is the affectation of tranquility; just as the lack of feeling and moral strength is accustomed to disguise itself as incisive coldness and detachment." [Nietzsche (1874) [Breazeale (1983), 93]] Nietzsche thus drew our attention to a deep problem: the problem of the inherent instability of both objectivity and the scientific self that practiced it. We have seen that the proposal of adopting *structural objectivity* as a solution to this problem was not successful. The program of the *Aufbau* could not be pulled through. A pluralism of various different language forms instead took its place as the guiding principle in the Carnapian theory of science. As Mormann puts it, "[f]rom *Logical Syntax* onwards, Carnap no longer was content to rationally reconstruct the world of scientific knowledge in a neat and orderly manner, rather, he aimed at the logical conquest of the entire universe of possible worlds". [*ibid.*] In fact, Mormann sees a deep analogy between the programs of new philosophy provided by Nietzsche and Carnap. The following passages from Nietzsche remind us very lively of the metaphors devised in the *Syntax*:

> Get on the Ships! — ... [We need] ... new philosophers! The moral earth, too, is round! The moral earth, too, has its antipodes! The antipodes, too, have their right to exist! There is yet another world to be discovered — and more than one! On the ships, you philosophers! [*The Gay Science*, Book IV, §289]

> ... finally the horizon seems clear again, even if not bright; finally our ships may set out again, set out to face any danger; every daring of the lover of knowledge is allowed again; the sea, our sea, lies open again; maybe there has never been such an "open sea". [*ibid*,, Book V, §343]

In his *Carnap and Twentieth-Century Thought* Carus wholeheartedly endorses this romantic reading of Carnapian pluralism, treating (along Thomas Mormann) Carnap's philosophical program as a science of possibilities [*Möglichkeitswissenschaft*]. But the romantic roots of Carnap's pluralism went deeper than the merely conceptual or purely formal ones: the strict dichotomy between the regions of *Geist* and *Leben* that in such dramatic a way characterized German thought at the turn of the twentieth century, also hindered Carnap from fully making evident the kernel of Enlightenment thought in his philosophy. Carnap was led to strict noncognitivism with respect to values and value judgements (as far as his published writings may be taken to represent his final views on these matters). From 1928 onwards, his fundamental values belonged and were nourished by the region of *Leben* and did not thus enter the region of rational deliberation. This fundamental dualism in his thought attests to the fact that Carnap's Enlightenment was not practical, and thus definitely not 'Radical' or spinozist; Carnap's Enlightenment was restricted to the purely theoretical. Nevertheless, there remains a strange feeling of an unreleased tension in his philosophical program. In the light of the history of objectivity and the tradition of intellectual virtues, including the flirt with the ideas of Nietzsche, Carnap seemed to have all the necessary tools in hand to try out to construct a unique synthesis out of these elements, vindicating the time-honored tradition of intellectual virtues and the techniques of the scientific self. Carnap, however, never took to heart the possibility to deliberate upon these matters in a rational way. Still, these very same problems haunt us as today as vividly as they did in the heyday of logical empiricism, and we may safely predict that they will continue to do so in the future.

In a summary, despite its shortcomings, Carus' interpretation of Carnap's philosophical work and its underlying motivation seem to be on the right track, as far as its cultural background is concerned. It is backed up by ample textual and contextual evidence, and more importantly, it enables us to formulate problems that Carnap never ventured to formulate explicitly but which he nonetheless found important. The import of Carus' interpretation is that he has stated Carnap's method in a language that is familiar to us and congruent with our contemporary problems. Moreover, making Carnap's approach in tackling those problems part of our own quest of structuring our knowledge and practice, we have incorporated tools into our thinking that facilitate our struggle to answer the most important questions of life and society more coherently and purposefully than would be possible without them. But as we have seen, the straightforward reading of Carnap's program as a modern form of Enlightenment does not hold water. Additional research is necessary to provide a balanced picture of Carnap's development and his mature program.

In the end, however, a humble attitude towards these questions prevails. This is in many ways an incomplete work, sketching only some of the background necessary for fully appreciating Carnap's contributions to philosophy, and even to the methods of analysis. Further work is necessary to enquire in more depth the specific technical problems pertaining to Carnap's overall program. Only through such enquiries can the full import of the ideal of explication be realized. And still, many questions that concern Carnap as a person remain in the dark, as important as they would be in forming a more adequate conception of the intimate relations between his scientific thought and moral convictions, providing a broader panorama of the *integrity* of his philosophy. I started my work with a quotation from Emil Staiger, emphasizing the importance of being open to the insights that history can provide in our intellectual undertakings. It is equally apt to finish this work with the following words of his:

Man is not an object about which wrong or even final pronouncements can be made. Man's essence is formed, arises in what he thinks of himself, in the unfolding of his self-consciousness. By giving specific answers to the question: "What is man?" we commit ourselves to specific possibilities. We become aware of ourselves, become conscious of ourselves in a specific way. Thus, one can say: In every system, in the worldview of every writer something of what man is capable of being is realized. The truth of such a worldview cannot be measured by what man actually is in the depth of his being. For this actual person, this person in and of itself, does not exist. Or he exists only for a spirit we would have to call divine. Here, too, the truth can only be measured by the degree to which it can be fruitful, to which it is capable of elucidating our present and our past.

– Emil Staiger, *Grundbegriffe der Poetik* (1946)

Appendix A

ARCHIVAL SOURCES

The archival sources are cited in the text giving the date of composition with the following identifier:

WKA
Wiener Kreis Archiv, Rijksarchief in Noord-Holland, Haarlem, Netherlands.

Bibliography

[1] d'Alembert, Jean le Rond (1759) [1965]: *Essai sur les Elements de Philosophie*, edité avec une introduction par Richard N. Schwab (Hildesheim: Georg Olms Verlagsbuchhandlung).

[2] Almond, Brenda (1998): "Applied ethics: a normative view" in E. Morscher et al. (eds.) (1998), 273–284.

[3] Allport, Gordon W. (1941): *The Use of Personal Documents in Psychological Science* (New York: Social Science Research Council).

[4] Ash, Mitchell G. (1998): *Gestalt Psychology in German Culture, 1890–1967* (Cambridge: Cambridge University Press).

[5] Awodey, Steve & Carus, André (2001): "Carnap, Completeness and Categoricity: The *Gabelbarkeitssatz* of 1928", *Erkenntnis* **54**, 145–172. [A draft version published in (1999) as Technical Report No. CMU-PHIL-92, Carnegie Mellon University]

[6] Awodey, Steve & Carus, André (2004): "How Carnap Could Have Replied to Gödel" in S. Awodey & C. Klein (eds.) (2004), 203–223.

[7] Awodey, Steve & Carus, André (2007): "Carnap's Dream", *Synthese* **159 (1)**, 23–45. [A draft version published in (2003) as Technical Report No. CMU-PHIL-145, Carnegie Mellon University]

[8] Awodey, Steve & Klein, Carsten (eds.) (2004): *Carnap Brought Home – The View from Jena* (La Salle, Illinois: Open Court).

[9] Awodey, Steve & Reck, Erich H. (2002a): "Completeness and Categoricity, part I: Nineteenth-century Axiomatics to Twentieth-century Metalogic", *History and Philosophy of Logic* **23 (1)**, 1–30.

[10] Awodey, Steve & Reck, Erich H. (2002b): "Completeness and Categoricity, part II: Twentieth-century Metalogic to Twenty-first-century Semantics", *History and Philosophy of Logic* **23 (2)**, 77–94.

[11] Ayer, Alfred Jules (ed.) (1959): *Logical Positivism* (New York: The Free Press).

[12] Ayer, Alfred Jules (1968): *The Origins of Pragmatism* (London: Macmillan).

[13] Ayer, Alfred Jules (1971): *Russell and Moore — The Analytic Heritage* (London: Macmillan).

[14] Baker, Gordon: *Wittgenstein, Frege and the Vienna Circle* (Oxford: Basil Blackwell).

[15] Baldwin, Thomas (2001): *Contemporary Philosophy: Philosophy in English since 1945* (Oxford: Oxford University Press).

[16] Baldwin, Thomas (2003): *The Cambridge History of Philosophy 1870–1945* (Cambridge: Cambridge University Press).

[17] Bar-Hillel, Y.; Poznanski, E.I.J.; Rabin, M.O. & Robinson, A. (eds.) (1966): *Essays on the Foundations of Mathematics* (Jerusalem: the Magnes Press, The Hebrew University).

[18] Barrow-Green, June (1997): *Poincaré and the Three Body Problem* (Rhode Island: American mathematical society).

[19] Barwise, Jon (1989): *The Situation in Logic* (Stanford: CSLI).

[20] Beaney, Michael (2003): "Analysis", in *The Stanford Encyclopedia of Philosophy*, `http://www.plato.stanford.edu/entries/analysis`.

[21] Beaney, Michael (2004): "Carnap's Conception of Explication: From Frege to Husserl?" in S. Awodey & C. Klein (eds.) (2004), 117–150.

[22] Beaney, Michael (ed.) (2007a): *The Analytic Turn* (New York: Routledge).

[23] Beaney, Michael (2007b): "The analytic turn in early twentieth-century philosophy" in M. Beaney (ed.) (2007), 1–30.

[24] Beiser, Frederick C. (1993): *The Cambridge Companion to Hegel* (Cambridge: Cambridge University Press).

[25] Beiser, Frederick C. (2002): *German Idealism — The Struggle Against Subjectivism* (Cambridge, Massachusetts: Harvard University Press).

[26] Beiser, Frederick C. (2005): *Hegel* (New York: Routledge).

[27] Bentham, Jeremy (1843): "Essay on Logic" in *The Works of Jeremy Bentham*, edited by J. Bowring (Edinburgh). Vol. 8, 213–293.

[28] Bell, Daniel (1976): *The Cultural Contradictions of Capitalism* (London: Heinemann).

[29] Bell, D. & Cooper, N. (eds.) (1990): *The Analytic Tradition* (Oxford: Blackwell).

[30] Bennett, Jonathan (1971): *Locke, Berkeley, Hume — Central Themes* (Oxford: Clarendon Press).

[31] Bennett, Jonathan (2001): *Learning from Six Philosophers* — DESCARTES, SPINOZA, LEIBNIZ, LOCKE, BERKELEY, HUME (Oxford: Clarendon Press).

[32] Bereiter, Carl & Scardamalia, Marlene (1993): *Surpassing Ourselves — An Inquiry into the Nature and Implications of Expertise* (Chicago, Illinois: Open Court).

[33] Berkvens-Stevelinck, C.; Israel, J. & Posthumus Meyjes, G. H. M. (eds.) (1997): *The Emergence of Tolerance in the Dutch Repulic* (Leiden: Koninklijke Brill).

[34] Berlin, Isaiah (1956): *The Age of Enlightenment* (ed.) (New York: Mentor Books).

[35] Berlin, Isaiah (1960): "The Concept of Scientifc History", originally published in *History and Theory*, Vol. 1, No. 1, 1960, pp. 1–31, under the title "History and Theory: The Concept of Scientific History". Reprinted in W.H. Dray (ed.) (1966).

[36] Berlin, Isaiah (1980): *Concepts and Categories — Philosophical Essays*, edited by Henry Hardy, (Oxford: Oxford University Press).

[37] Berlin, Isaiah (1981): *Personal Impressions*, edited by Henry Hardy, (London: The Hogarth Press).

[38] Berlin, Isaiah & Jahanbegloo, Ramin (1991): *Conversations with Isaiah Berlin* (New York: Charles Scribner's Sons).

[39] Beth, Evert W. (1953): "On Padoa's Method in the Theory of Definition", *Indag. Math.* **15**, 330–339.

[40] Beth, Evert W. & Piaget, Jean (1966): *Mathematical Epistemology and Psychology* (Dordrecht: D. Reidel).

[41] Biletzki, A. & Matar, A. (eds.) (1998): *The Story of Analytic Philosophy* (London: Routledge).

[42] Bird, Graham (1972): *Philosophical Tasks — An introduction to some aims and methods in recent philosophy* (London: Hutchison University Library).

[43] Bird, Graham H. (ed.) (2006a): *A Companion to Kant* (Oxford: Blackwell).

[44] Bird, Graham H. (2006b): "Kant's Analytic Apparatus" in G. H. Bird (ed.) (2006a), 125–139.

[45] Bird, Graham H. (2006c): *Revolutionary Kant — A Commentary on the* CRITIQUE OF PURE REASON (Chicago, Illinois: Open Court).

[46] Birkhoff, Garrett (ed.) (1973): *A Source Book in Classical Analysis* (Cambridge, Massachusetts: Harvard University Press).

[47] Birkhoff, George David (1923): *Relativity and Modern Physics* (Cambridge, Massachusetts: Harvard University Press).

[48] Birkhoff, George David (1927): *Dynamical Systems* (New York: American Mathematical Society).

[49] Boi, L.; Flament, D. & Salanskis, J.-M. (eds.) (1992): *1830–1930: A Century of Geometry* (Berlin: Springer-Verlag).

[50] Bolzano, Bernard (1837) [1963]: *Grundlegung der Logik (Wissenschaftslehre I/II)*, ausgewählte Paragraphen aus der *Wissenschaftslehre, Band I und II*, mit ergänzenden Textzusammnefassungen einer Einleitung und Registern herausgegeben von Friedrich Kambartel, (Hamburg: Verlag von Felix Meiner).

[51] Bonnay, Dennis (2008): "Logicality and Invariance", *The Bulletin of Symbolic Logic*, Volume **14**, Issue 1 (2008), 29-68.

[52] Bordieau, Pierre (1975): "The Specificity of the Scientific Field and the Social Conditions of the Progress of Reason", *Social Science Information*, **14**, 1975

[53] Bos, Henk J. M. (2001): *Redefining Geometrical Exactness – Descartes' Transformation of the Early Modern Concept of Construction* (New York: Springer-Verlag).

[54] Braithwaite, R. (1953): *Scientific Explanation* (Cambridge: Cambridge University Press).

[55] Brandom, Robert B. (1994): *Making It Explicit* (Cambridge, Massachusetts: Harvard University Press).

[56] Brandom, Robert B. (2000): *Articulating Reasons* (Cambridge, Massachusetts: Harvard University Press).

[57] Breazeale, Daniel (1983) (ed.): *Nietzsche: Untimely Meditations* (Cambridge: Cambridge University Press).

[58] Breazeale, Daniel (1988) (ed.): *Fichte: Early Philosophical Writings* (Ithaca: Cornell University Press).

[59] Bruford, W. H. (1975): *The German Tradition of Self-Cultivation* (Cambridge: Cambridge University Press).

[60] Bunge, Mario (1979): *Scientific Research. 1. The Search for System* (Berlin: Springer).

[61] Butterfield, Jeremy & Earman, John (eds.) (2007): *Philosophy of Physics*, Parts A & B (Amsterdam: Elsevier).

[62] Carnap, Rudolf (1922): *Der Raum: Ein Beitrag zur Wissenschaftslehre*, KANT-STUDIEN, Ergänzungshefte, Nr. **56** (Berlin: Reuther & Reichard).

[63] Carnap, Rudolf (1923): "Über die Aufgabe der Physik und die Anwendung des Grundsatzes der Einfachstheit", *Kant-Studien* **28**, 90–107.

[64] Carnap, Rudolf (1924): "Dreidimensionalität des Raumes und Kausalität: Eine Untersuchung über den logischen Zusammenhang zweier Fiktionen", *Annalen der Philosophie und philosophischen Kritik* **4**, 105–130.

[65] Carnap, Rudolf (1928a) [1967]: *Der Logische Aufbau der Welt* in *The Logical Structure of the World & Pseudoproblems in Philosophy*, translated by Rolf A. George (Berkeley: University of Califonia Press).

[66] Carnap, Rudolf (1928b) [1967]: *Scheinprobleme in der Philosophie* in *The Logical Structure of the World & Pseudoproblems in Philosophy*, translated by Rolf A. George (Berkeley: University of Califonia Press).

[67] Carnap, Rudolf (1929): *Abriss der Logistik – Mit Besonderer Berücksichtigung der Relationsteorie und ihrer Anwendungen* (Wien: Verlag von Julius Springer).

[68] Carnap, Rudolf (1930a): "Die alte und die neue Logik", *Erkenntnis* **1**, 12–26.

[69] Carnap, Rudolf (1930b): "Bericht über Untersuchungen zur allgemeinen Axiomatik", *Erkenntnis* **1**, 303–307.

[70] Carnap, Rudolf (1932a): "Die physikalische Sprache als Universalsprache der Wissenschaft", *Erkenntnis* **2**, 432–465.

[71] Carnap, Rudolf (1932b): "Psychologie in physikalischer Sprache", *Erkenntnis* **3**, 107–142.

[72] Carnap, Rudolf (1932c): "Erwiderung auf die vorstehenden Aufsätze von E. Zilsel und K. Duncker", *Erkenntnis* **3**, 177–188.

[73] Carnap, Rudolf (1932d): "Über Protokollsätze", *Erkenntnis* **3**, 215–228.

[74] Carnap, Rudolf (1932e): "Russell und Whitehead, Einführung in die mathematische Logik [Besprechung]", *Erkenntnis* **3**, 436–437.

[75] Carnap, Rudolf (1934a) [1937]: *Logische Syntax der Sprache* (Wien: Springer), translated by Amethe Smeaton as *The Logical Syntax of Language* (London: Routledge).

[76] Carnap, Rudolf (1934b) [1987]: *Die Aufgabe der Wissenschaftslogik* (Wien: Gerold), translated by H. Kaal as "The Task of the Logic of Science", in B. McGuinness (ed.) (1987), 46–66.

[77] Carnap, Rudolf (1935a): "Formalwissenschaft und Realwissenschaft [Bericht (Prag)]", *Erkenntnis* **5**, 30–37.

[78] Carnap, Rudolf (1935b): "Quine, A System of Logistic [Besprechung]", *Erkenntnis* **5**, 285–287.

[79] Carnap, Rudolf (1935c): "Dubislaw, Naturphilosophie [Besprechung]", *Erkenntnis* **5**, 287–288.

[80] Carnap, Rudolf (1935d): "Heyting, Mathematische Grundlagenforschung, Intuitionismus, Beweistheorie [Besprechung]", *Erkenntnis* **5**, 288–289.

[81] Carnap, Rudolf (1935e): "Popper, Logik der Forschung [Besprechung]", *Erkenntnis* **5**, 290–294.

[82] Carnap, Rudolf (1936–37) [1950]: *Testability and Meaning*, reprinted, with the permission of the editor and the publishers, The Williams & Wilkins Co., Baltimore, from Philosophy of Science, Vol. III, 1936 , and Vol. IV, 1937; with corrigenda and additional bibliography by Prof. Carnap (New Haven, Connecticut: Graduate Philosophy Club, Yale University).

[83] Carnap, Rudolf (1938) [1955]: *Logical Foundations of the Unity of Science* (Chicago: The University of Chicago Press).

[84] Carnap, Rudolf (1939) [1955]: *Foundations of Logic and Mathematics* (Chicago: The University of Chicago Press).

[85] Carnap, Rudolf (1942) [1959]: *Introduction to Semantics* in *Introduction to Semantics* AND *Formalization of Logic (Two Volumes in One)* (Cambridge, Massachusetts: Harvard University Press).

[86] Carnap, Rudolf (1943) [1959]: *Formalization of Logic* in *Introduction to Semantics* AND *Formalization of Logic (Two Volumes in One)* (Cambridge, Massachusetts: Harvard University Press).

[87] Carnap, Rudolf (1947) [1956]: *Meaning and Necessity*, enlarged edition (Chicago: The University of Chicago Press).

[88] Carnap, Rudolf (1950a) [1956]: "Empiricism, Semantics, and Ontology", reprinted in *Meaning and Necessity*, 2nd edition (Chicago: The University of Chicago Press), 205–221.

[89] Carnap, Rudolf (1950b) [1956]: *The Logical Foundations of Probability* (Chicago: The University of Chicago Press).

[90] Carnap, Rudolf (1952a): *The Continuum of Inductive Methods* (Chicago: The University of Chicago Press).

[91] Carnap, Rudolf (1952b) [1956]: "Meaning Postulates", reprinted in *Meaning and Necessity*, 2nd edition (Chicago: The University of Chicago Press), 222–229.

[92] Carnap, Rudolf (1954) [1956]: "On Belief Sentences. Reply to Alonzo Church.", reprinted in *Meaning and Necessity*, 2nd edition (Chicago: The University of Chicago Press), 230–232.

[93] Carnap, Rudolf (1955a) [1956]: "Meaning and Synonymy in Natural Languages", reprinted in *Meaning and Necessity*, 2nd edition (Chicago: The University of Chicago Press), 233–247.

[94] Carnap, Rudolf (1955b) [1956]: "On Some Concepts of Pragmatics", reprinted in *Meaning and Necessity*, 2nd edition (Chicago: The University of Chicago Press), 248–250.

[95] Carnap, Rudolf (1956): "The Methodological Character of Theoretical Concepts" in H. Feigl & M. Scriven (eds.) (1956), 38–76.

[96] Carnap, Rudolf (1963): "Intellectual Autobiography" and "The Philosopher Replies" in P. Schilpp (ed.) (1963), 3–84, 859–1013.

[97] Carnap, Rudolf (1966a): *Philosophical Foundations of Physics: An Introduction to the Philosophy of Science* (New York: Basic Books).

[98] Carnap, Rudolf (1966b): "On the use of Hilbert's ϵ-operator in scientific theories" in Y. Bar-Hillel, E.I.J. Poznanski, M.O. Rabin, & A. Robinson (eds.) (1966), 156–164.

[99] Carnap, Rudolf (1968a): "Inductive Logic and Inductive Intuition" in I. Lakatos (ed.) (1968), 258–267.

[100] Carnap, Rudolf (1968b): "On rules of acceptance [Discussion]" in I. Lakatos (ed.) (1968), 146–149.

[101] Carnap, Rudolf (1968c): "The concept of constituent-structure [Discussion]" in I. Lakatos (ed.) (1968), 218–219.

[102] Carnap, Rudolf (1971a): "Inductive Logic and Rational Decisions" in R. Carnap & R. Jeffrey (eds.) (1971), 5–32.

[103] Carnap, Rudolf (1971b): "A Basic System of Inductive Logic, Part I" in R. Carnap & R. Jeffrey (eds.) (1971), 33–166.

[104] Carnap, Rudolf (1993): *Mein Weg in die Philosophie* (Stuttgart: Philipp Reclam jun.).

[105] Carnap, Rudolf (2000) [1927–1930]: *Untersuchungen zur allgemeinen Axiomatik*, herausgegeben von Thomas Bonk & Jesus Mosterin (Darmstadt: Wissenschaftliche Buchgesellschaft).

[106] Carnap, Rudolf & Stegmüller, Wolfgang (1959): *Induktive Logik und Wahrscheinlichkeit* (Wien: Springer-Verlag).

[107] Carnap, Rudolf & Jeffrey, Richard (eds.) (1971): *Studies in Inductive Probability*. Vol. I. (Berkeley, California: University of California Press).

[108] Carus, André W. (2002): "The Philosopher without Qualities" in M. Heidelberger & F. Stadler (eds.) (2002), 369–377.

[109] Carus, André W. (2007): *Carnap and Twentieth-Century Thought – Explication as Enlightenment*. (Cambridge: Cambridge University Press).

[110] Cassidy, David C. (2000): "A Historical Perspective on *Copenhagen*" *Physics Today*, July 2000, 28–32.

[111] Cassirer, Ernst (1923): *Substanzbegriff und Funktionsbegriff* (Berlin: Verlag von Bruno Cassirer).

[112] Cassirer, Ernst (1950): *The Problem of Knowledge – Philosophy, Science and History since Hegel*, trans. William H. Woglom & Charles W. Hendel (New Haven: Yale University Press).

[113] Cassirer, Ernst (1951): *The Philosophy of the Enlightenment*, trans. Fritz C. A. Koelin & James P. Pettegrove (Princeton: Princeton University Press).

[114] Cohen, I. Bernard & Smith, George E. (eds.) (2002): *The Cambridge Companion to Newton* (Cambridge: Cambridge University Press).

[115] Cohen, Jonathan L. (1986): *The Dialogue of Reason — An Analysis of Analytical Philosophy* (Oxford: Clarendon Press).

[116] Collingwood, Robin George (1936): *Human Nature and Human History*. Proceedings of the British Academy, Volume XXII. (London: Humphrey Milford Amen House).

[117] Collingwood, Robin George (1940): *An Essay on Metaphysics* (Oxford at the Clarendon Press).

[118] Collingwood, Robin George (1965): *Essays in the Philosophy of History*, edited by W. Debbins, (Austin: University of Texas Press).

[119] Collingwood, Robin George (1978): *An Autobiography* (Oxford: Clarendon Press).

[120] Collingwood, Robin George (1993): *The Idea of History*, Second edition (Oxford: Clarendon Press).

[121] Collingwood, Robin George (2005a): *An Essay on Philosophical Method* (Oxford: Clarendon Press).

[122] Collingwood, Robin George (2005b): *The Philosophy of Enchantment* (Oxford: Clarendon Press).

[123] Comte, Auguste (1830–1842) [(1974)]: *The Positive Philosophy*, freely translated and condensed by Harriet Martineau (New York: AMS Press, Inc.).

[124] Connelly, James (2006): "Postmodern Skepticism, Truth and History" in A. L. Macfie (ed.) (2006).

[125] Connes, Alain; Lichnerowicz, André; Schützenberger, Marcel Paul (2001): *Triangle of Thoughts*, translated by Jennifer Gage (Providence, Rhode Island: American Mathematical Society).

[126] Cushing, James T. (1998): *Philosophical Concepts in Physics — The Historical Relation between Philosophy and Scientific Theories* (Cambridge: Cambridge University Press).

[127] van Dalen, Dirk (1999): *Mystic, Geometer and Intuitionist — The Life of L.E.J. Brouwer* (Oxford: Clarendon Press).

[128] Damasio, Antonio R. (2003): *Looking for Spinoza: Joy, Sorrow and the Feeling Brain* (London: Houghton Mifflin Harcourt).

[129] Damerow, Peter; Freudenthal, Gideon; McLaughlin, Peter & Renn, Jürgen (2004): *Exploring the Limits of Preclassical Mechanics* (Berlin: Springer).

[130] Danto, Arthur C. (1965): *Analytical philosophy of history* (London: Cambridge University Press).

[131] Danto, Arthur C. (1985): *Narration and Knowledge* (New York: Columbia University Press).

[132] Daston, Lorraine & Galison, Peter (2007): *Objectivity* (New York: Zone Books).

[133] de Jonge, Eccy (2006): "The Place of History in Spinoza's Metaphysics" in A. L. Macfie (ed.) (2006).

[134] Dingler, Hugo (1919): *Die Grundlagen der Physik — Synthetische Prinzipien der mathematischen Naturphilosophie* (Berlin: Vereinigung wissenschaftlicher Verleger Walter de Gruyter & Co.).

[135] Dray, William, H. (ed.) (1966): *Philosophical Analysis and History* (New York: Harper & Row, Publishers).

[136] Dummett, Michael (1991): *The Logical Basis of Metaphysics* (Cambridge, Massachusetts: Harvard University Press).

[137] Dummett, Michael (1993): *Origins of Analytic Philosophy* (London: Duckworth)

[138] Earman, John (1986): *Primer on Determinism* (Dordrecht: D. Reidel).

[139] Earman, John (1989): *World Enough and Space-Time — Absolute versus Relational Theories of Space and Time* (Cambridge, Massachusetts: The MIT Press).

[140] Earman, John (ed.) (1992): *Inference, Explanation, and Other Frustrations* (Berkeley: University of California Press).

[141] Earman, John (2007): "Aspects of Determinism in Modern Physics" in J. Butterfield & J. Earman (eds.) (2007), Part B.

[142] Eddington, Arthur Stanley (1920): *Space, Time and Gravitation* (Cambridge: Cambridge University Press).

[143] Feferman, S.; Dawson, J. W.; Kleene, S. C.; Moore, G. H.; Solovay, R. M. & van Heijenoort, J. (eds.) (1986): *Kurt Gödel: Collected Works*, Volume I (New York/Oxford: Oxford University Press/Clarendon Press).

[144] Feferman, S.; Dawson, J. W.; Goldfarb W.; Parsons, C. & Solovay, R. M. (eds.) (1995): *Kurt Gödel: Collected Works*, Volume III (Oxford: Oxford University Press).

[145] Feferman, S.; Dawson, J. W.; Goldfarb W.; Parsons, C. & Sieg, W. (eds.) (2003): *Kurt Gödel: Collected Works*, Volume IV: Correspondence A–G (Oxford: Clarendon Press).

[146] Feigl, Herbert & Scriven, Michael (eds.) (1956): *Minnesota Studies in the Philosophy of Science. Volume I — The Foundations of Science and the Concepts of Psychology and Psychoanalysis* (Minneapolis, Minnesota: University of Minnesota Press).

[147] Foucault, Michel (1989): "Subjectivité et vérité", *Résumé de cours, 1970–82* (Paris: Julliard).

[148] Forster, Michael (1993): "Hegel's dialectical method" in Frederick Beiser (ed). (1993), 130–170.

[149] Francis, Mark (ed.) (1985): *The Viennese Enlightenment* (London: Croom Helm).

[150] Fraenkel, Adolf (1928) [1919/1923]: *Einleitung in die Mengenlehre*, revised 2nd and 3rd editions (Berlin: Springer Verlag).

[151] Frege, Gottlob (1879): *Begriffschrift — Eine der Arithmetischen nachgebildete Formelsprache des reinen Denkens* (Halle: Nebert).

[152] Frege, Gottlob (1893/1903): *Grundgesetze der Arithmetik*, Volumes 1–2 (Jena: Pohle).

[153] Frege, Gottlob (1910–1914) [2004]: *Frege's Lectures on Logic — Carnap's Student Notes, 1910–1914*, E. Reck & S. Awodey (ed. and trans.) (La Salle, Illinois: Open Court).

[154] Friedman, Michael (1999): *Reconsidering Logical Positivism* (Cambridge: Cambridge University Press).

[155] Friedman, Michael (2002): "Kant, Kuhn, and the Rationality of Science" in M. Heidelberger & F. Stadler (eds.) (2002), 25–41.

[156] Friedman, Michael & Creath, Richard (eds.) (2007): *The Cambridge Companion to Carnap* (Cambridge: Cambridge University Press).

[157] Fulbrook, Mary (2002): *Historical Theory* (London: Routledge).

[158] Fulbrook, Mary (2006): "Why All Historical Accounts Are Inevitably Theoretical; but Why Some Accounts Are Preferable to Others" in A. L. Macfie (ed.) (2006).

[159] Gardiner, Patrick (1952): *The Nature of Historical Explanation* (Oxford: Oxford University Press).

[160] Gardiner, Patrick (1995): "Historicism" in *The Oxford Companion to Philosophy* (Oxford: Oxford University Press).

[161] Gaukroger, Stephen (2006): *The Emergence of a Scientific Culture. Science and the Shaping of Modernity, 1210–1685* (Oxford: Clarendon Press).

[162] Geier, Manfred (1992): *Der Wiener Kreis* (Hamburg: Rowohlt).

[163] Gell-Mann, Murray & Lloyd, Seth (1996): "Information Measures, Effective Complexity and Total Information" in *Complexity*, Volume **2**, Issue **1**, 44–52.

[164] Glock, Hans-Johann (1997): *The Rise of Analytic Philosophy* (Oxford: Blackwell).

[165] Glock, Hans-Johann (2008): *What Is Analytic Philosophy?* (Cambridge: Cambridge University Press).

[166] Gödel, Kurt (1931): "Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I", *Monatshäfte für Matematik und Physik* **38**, 173–198. Translated as "On formally undecidable propositions of *Principia Mathematica* and related systems I" in S. Feferman et al. (eds.) (1986), 144–195.

[167] Goethe, Johann Wolfgang (1979): *Farbenlehre*, edited by Gerhard Ott & Heinrich O. Proskauer, 5 volumes (Stuttgart: Verlag Freies Geistesleben).

[168] Goldfarb, Warren (1995): "Introductory Note to [Gödel (1995)]" in S. Feferman et al. (eds.) (1995), 324–334.

[169] Goodman, Nelson (1977): *The Structure of Appearance*, Third Edition with an Introduction by Geoffrey Hellman (Dordrecht: D. Reidel).

[170] Goodman, Nelson (1990): *A Study of Qualities*, academic dissertation, originally submitted in 1940, (New York: Garland Publishing).

[171] Gorman, Jonathan (2007): *Historical Judgement — The Limits of Historiographical Choice* (Stocksfield: Acumen).

[172] Grattan-Guinnes, Ivor (1970): *The Development of the Foundations of Mathematical Analysis from Euler to Riemann* (Cambridge, Massachusetts: MIT Press).

[173] Grattan-Guinnes, Ivor (2000): *The Search for Mathematical Roots, 1870–1940* (Princeton: Princeton University Press).

[174] Griffin, Nicholas (ed.) (2003): *The Cambridge Companion to Bertrand Russell* (Cambridge: Cambridge University Press).

[175] Grünbaum, Adolf (1971): "The Meaning of Time" in *Basic Issues in the Philosophy of Time*. Edited by Freeman & Sellars. La Salle: Open Court, 195–228.

[176] Guyer, Paul (ed.) (2006): *The Cambridge Companion to Kant and Modern Philosophy* (Cambridge: Cambridge University Press).

[177] Habermas, Jürgen (2001): "Learning from Catastrophe? A Look Back at the Short Twentieth Century." in *The Postnational Constellation: Political Essays*. (Cambridge, Massachusetts: MIT Press).

[178] Hacker, P. M. S. (1996): *Wittgenstein's Place in Twentieth Century Analytic Philosophy* (Oxford: Blackwell).

[179] Hacking, Ian (1965): *Logic of Statistical Inference* (Cambridge: Cambridge University Press).

[180] Hacking, Ian (1983): *Representing and Intervening — Introductory topics in the philosophy of natural science* (Cambridge: Cambridge University Press).

[181] Hacking, Ian (1990): *The Taming of Chance* (Cambridge: Cambridge University Press).

[182] Hacking, Ian (2000): *Historical Ontology* (Cambridge, Massachusetts: Harvard University Press).

[183] Haddock, Guillermo E. Rosado (2008): *The Young Carnap's Unknown Master* (Hampshire: Ashgate).

[184] Hager, Paul (1994): *Continuity and Change in the Development of Russell's Philosophy* (Dordrecht: Kluwer Academic Publishers).

[185] Hager, Paul (2003): "Russell's Method of Analysis" in N. Griffin (ed.) (2003), 310–331.

[186] Hanna, R. (2001): *Kant and the Analytic Tradition* (Oxford: Oxford University Press).

[187] Harper, William (2002): "Newton's argument for universal gravitation" in I. Bernard Cohen & George E. Smith (eds.) (2002), 174–201.

[188] Hegel, G. W. F. (1944): *The Philosophy of History*, trans. J. Sibree (New York: Willey Book Co.)

[189] Heidelberger, Michael & Stadler, Friedrich (eds.)(2002): *History of Philosophy of Science — New trends and perspectives* (Dordrecht: Kluwer Academic Publishers).

[190] Hempel, Carl G. (1962): "Explanation in Science and in History", originally published in *Frontiers of Science and Philosophy*, ed. R. G. Colodny, Pittsburgh, 1962, pp. 9–33; reprinted in W.H. Dray (ed.) (1966).

[191] Hintikka, Jaakko (1973): *Logic, Language-Games and Information* (Oxford at the Clarendon Press).

[192] Hintikka, Jaakko (1975a): *The Intentions of Intentionality and Other New Models for Modalities* (Dordrecht: D. Reidel).

[193] Hintikka, Jaakko (ed.) (1975b): *Rudolf Carnap, Logical Empiricist — Materials and Perspectives* (Dordrecht: D. Reidel).

[194] Hintikka, Jaakko (1992): "The Concept of Induction in the Light of the Interrogative Approach to Inquiry" in John Earman (ed.) (1992), 23–43.

[195] Hintikka, Jaakko (1996): *Ludwig Wittgenstein — Half Truths and One-and-a-Half-Truths*, Selected Papers I (Dordrecht: Kluwer Academic Publishers).

[196] Hintikka, Jaakko (1997): *Lingua Universalis vs. Calculus Ratiocinator: An Ultimate Presupposition of Twentieth-century Philosophy*, Selected Papers II (Dordrecht: Kluwer Academic Publishers).

[197] Hintikka, Jaakko (1998): *Language, Truth and Logic in Mathematics*, Selected Papers III (Dordrecht: Kluwer Academic Publishers).

[198] Hintikka, Jaakko & Remes, Unto (1974): *The Method of Analysis: Its Geometrical Origin and Its General Significance*, No. **25** in *Boston Studies in the Philosophy of Science* (Dordrecht: D. Reidel).

[199] Hintikka, Jaakko & Suppes, Patrick (1966): *Aspects of Inductive Logic* (Dordrecht: D. Reidel).

[200] Hintikka, Jaakko & Suppes, Patrick (1970): *Information and Inference* (Dordrecht: D. Reidel).

[201] Hintikka, Merrill B. & Hintikka, Jaakko (1986): *Investigating Wittgenstein* (Oxford: Basil Blackwell).

[202] Hoefer, Carl (2010): "Causal Determinism", Stanford Encyclopedia of Philosophy (SEP). Retrieved March 22, 2010 from `http://plato.stanford.edu/entries/determinism-causal/`.

[203] Holmes, Philip (1990): "Poincaré, celestial mechanics, dynamical-systems theory and 'chaos', *Physics Reports* **193** (3), 137-163.

[204] Husserl, Edmund (1970) [1891]: *Philosophie der Arithmetik*, in *Gesammelte Werke*, XII (Den Haag: Martinus Nijhoff).

[205] Husserl, Edmund (2001) [1900–1901]: *Logical Investigations*, 2 vols, tr. J.N. Findlay, ed. D. Moran (London: Routledge).

[206] Hylton, Peter (1990): *Russell, Idealism and the Emergence of Analytic Philosophy* (Oxford: Clarendon Press).

[207] Israel, Jonathan I. (1999): *Locke, Spinoza and the Philosophical Debate Concerning Toleration in the Early Enlightenment (c. 1670 – c. 1750)* (Amsterdam: Koninklijke Nederlandse Akademie van Wetenschappen).

[208] Israel, Jonathan I. (2001): *Radical Enlightenment — Philosophy and the Making of Modernity 1650–1750* (Oxford: Oxford University Press).

[209] Israel, Jonathan I. (2006): *Enlightenment Contested — Philosophy, Modernity and the Emancipation of Man 1650–1752* (Oxford: Oxford University Press).

[210] Jacob, Margaret C. (1981/2003): *The Radical Enlightenment: Pantheists, Freemasons and Republicans*, (New Orleans: Cornerstone Book Publishers).

[211] James, William (1956): "The Dilemma of Determinism" in *The Will to Believe* (New York: Dover Publications).

[212] Janik, Allan (2001): *Wittgenstein's Vienna Revisited* (New Brunswick, New Jersey: Transaction Publishers).

[213] Janik, Allan & Toulmin, Stephen E. (1973): *Wittgenstein's Vienna* (Chicago: Ivan R. Dee).

[214] Jeffrey, Richard (ed.) (1980): *Studies in Inductive Probability*. Vol. II. (Berkeley, California: University of California Press).

[215] Kaila, Eino (1979): *Reality and Experience — Four Philosophical Essays*, edited by Robert S. Cohen, (Dodrecht: D. Reidel).

[216] Kalli, Pekka (1995): *Cassirer ja uuskantilaisuus*, academic dissertation, (Tampere: University of Tampere, Department of Philosophy).

[217] Kant, Immanuel (1970) [1784]: "Idea for a Universal History with a Cosmopolitan Purpose" in *Kant's Political Writings*, ed. Hans Reiss, 41–53. (Cambridge: Cambridge University Press).

[218] Kant, Immanuel (1998) [1789]: *Logik-Vorlesungen, Unveröffentliche Nachschriften I – Logik Bauch*, bearbeitet von Tillman Pinder (Hamburg: Felix Meiner Verlag).

[219] Keyserling, Arnold (1965): *Der Wiener Denkstil — Mach, Carnap, Wittgenstein* (Wien: Stiasny Verlag).

[220] Kluckhohn, C. & Kelly, W.H. (1945): "The Concept of Culture" in *The Science of Man in the World Crisis*, edited by R. Linton, 78–105 (New York: Columbia University Press).

[221] Kolb, Eberhard (1988): *The Weimar Republic*, trans. P. S. Falla (London: Unwin Hyman).

[222] Kolmogorov, A.N. (1984): "On logical foundations of probability theory" in K. Itô & J.V. Prokhorov (1984) (eds.): *Probability Theory and Mathematical Statistics. Lecture Notes in Mathematics*, **1021** (Berlin: Springer-Verlag).

[223] Korhonen, Anssi (2007): *Logic as the Universal Science – Bertrand Russell's Early Philosophy of Logic and Its Philosophical Context*, academic dissertation (Helsinki: University of Helsinki, Department of Philosophy).

[224] Kovalevskaya, Sonya E. (1978): *A Russian Childhood*, translated by Beatrice Stillman (New York: Springer-Verlag). First edition published in 1889 in Swedish as a novel *From Russian Life: the Rajevski Sisters*. First published in autobiographical form in Russian in 1890.

[225] Krauth, Lothar (1970): *Die Philosohpie Carnaps* (Wien: Springer Verlag).

[226] von Krockow, Christian Graf (1990): *Die Deutschen in Ihrem Jahrhundert* (Hamburg: Rowohlt).

[227] Krylov, Alekseĭ, N. (1933): *Leonhard Euler* (Moscow-Leningrad: Izdat. Akad. Nauk SSSR).

[228] Kusch, Martin (1989): *Language as Calculus vs. Language as Universal Medium — A Study in Husserl, Heidegger and Gadamer*, an academic dissertation, (Dordrecht: Kluwer Academic Publishers).

[229] Kyburg, Henry E. Jr. (1974): *The Logical Foundations of Statistical Inference* (Dordrecht: D. Reidel).

[230] Lakatos, Imre (ed.) (1968): *The Problem of Inductive Logic* (Amsterdam: North Holland).

[231] van Lambalgen, Michiel (1987): *Random Sequences*. Academisch Proefschrift, University of Amsterdam.

[232] van Lambalgen, Michiel (1989): "Algorithmic information theory", *Journal of Symbolic Logic* **54**, 1389–1400.

[233] Langford, Peter E. (2005): *Vygotsky's Developmental and Educational Psychology* (New York: Psychology Press).

[234] de Laplace, Pierre Simon (1820): *Théorie analytique des probabilités* (Paris: V. Courcier).

[235] Lee, Sir S. & Boas, Frederick Samuel (1929): *Elizabethan and Other Essays* (London: Ayer Publishing).

[236] Lloyd, Christopher (1993): *The Structures of History* (Oxford: Blackwell).

[237] Luria, A.R. & Vygotsky, L.S. (1992): *Ape, Primitive Man and Child* (London: Harvester Wheatsheaf).

[238] Macfie, Alexander Lyon (ed.) (2006): *The Philosophy of History* (Hampshire: Palgrave Macmillan).

[239] Mac Lane, Saunders (1938): "Carnap on Logical Syntax", *Bulletin of the American Mathematical Society*, 171–176.

[240] Majer, Ulrich (2002): "Hilbert's Program to Axiomatize Physics (in Analogy to Geometry) and its Impact on Schlick, Carnap and other Members of the Vienna Circle" in M. Heidelberger & F. Stadler (eds.) (2002), 213–224.

[241] Manninen, Juha (1995): "Otto Neurath über die Unmöglichkeit einer 'privaten Sprache' und die Möglichkeit der Visualisierung", in Hans Jörg Sandkühler (ed.), *Interaktionen zwischen Philosophie und empirischen Wissenschaften. Philosophie- und Wissenschaftsgeschichte zwischen Francis Bacon und Ernst Cassirer, Philosophie und Geschichte der Wissenschaften*. (Frankfurt am Main / Berlin / Bern / New York / Paris / Wien: Studien und Quellen, Peter Lang: Europäischer Verlag der Wissenschaften), 405–423.

[242] Manninen, Juha (2001): "Das verbotene Buch des Wiener Kreises", in Hans Heinz Holz and Domenico Losurdo (eds.), *Topos. Internationale Beiträge zur dialektischen Theorie **17**: Ideologie* (Napoli: Edizioni La Città del Sole), 65–77.

[243] Manninen, Juha (2002a): "Uuden filosofisen liikkeen ja sen manifestin synty" (The birth of a new philosophical movement and its manifesto), in Ilkka Niiniluoto and Heikki J. Koskinen (eds.) (2002), 27–128.

[244] Manninen, Juha (2002b): "Ennen Wienin piiriä: Moritz Schlick, Hans Reichenbach ja Eino Kaila" (Before the Vienna Circle: Moritz Schlick, Hans Reichenbach and Eino Kaila), in Lauri Mehtonen and Kari Väyrynen (eds.), *Järjen todellisuus. Juhlakirja Markku Mäelle*, Aate- ja oppihistorian tutkimuksia **26**, (Oulu: Sophopolis), 245–275.

[245] Manninen, Juha (2002c): "Eino Kaila ja tie Wienin piiriin" (Eino Kaila and his way to the Vienna Circle), in *Tieteessä tapahtuu* **2002:2**, 34–47.

[246] Manninen, Juha (2002d): "Realisti Wienin piirin väittelyissä" (Eino Kaila as a realist in the debates of the Vienna Circle), in *Tieteessä tapahtuu* **2002:7**, 40–52.

[247] Manninen, Juha (2002e): "Positivismin kriitikosta positivistiksi" (The transformation of Eino Kaila's views from a criticism of positivism to positivism), in *Tieteessä tapahtuu* **2002:8**, 31–41.

[248] Manninen, Juha (2003): "Towards a Physicalistic Attitude", in Friedrich Stadler (ed.), *The Vienna Circle and Logical Empiricism. Re-evaluation and Future Perspectives*, Vienna Circle Institute Yearbook **10**, (Dordrecht / Boston / London: Kluwer Academic Publishers), 133–150.

[249] Manninen, Juha (2004a): "Wien, Cambridge, Königsberg: filosofisia tapahtumia 1929-1930" (Philosophical occasions: Vienna, Cambridge, Königsberg 1929-1930), in *Tieteessä tapahtuu* **2004:4**, 17–26.

[250] Manninen, Juha (2004b): "Wittgenstein ja Wienin piiri" (Wittgenstein and the Vienna Circle), in *niin & näin. Filosofinen aikakauslehti* **02/2004**, 134–153.

[251] Manninen, Juha (2009a): "Välittömät, yksityiset elämykset Wienin piirin ja Wittgensteinin pulmana" (Private lived experiences as a problem for the Vienna Circle and Wittgenstein), in Juha Manninen and Risto Vilkko (eds.) (2009), 194–221.

[252] Manninen, Juha (2009b): "Between the Vienna Circle and Ludwig Wittgenstein. The Philosophical Teachers of Georg Henrik von Wright", in Juha Manninen and Friedrich Stadler (eds.), *The Vienna Circle in the Nordic Countries*, Vienna Circle Institute Yearbook, (Wien / New York: Springer).

[253] Manninen, Juha (forthcoming): *Developments and Tensions within the Vienna Circle*, Veröffentlichungen des Instituts Wiener Kreis (Wien / New York: Springer).

[254] Manninen, Juha & Stadler, Friedrich (eds.) (2009): *The Vienna Circle in the Nordic Countries*, Vienna Circle Institute Yearbook (Wien: Springer).

[255] Manninen, Juha & Vilkko, Risto (eds.) (2009): *Ajattelun välineet ja maailmat — Kirjoituksia Jaakko Hintikan filosofiasta* (Helsinki: Gaudeamus, Helsinki University Press).

[256] Martin, Rex (1977): *Historical Explanation — Re-enactment and Practical Inference* (Ithaca: Cornell University Press).

[257] Martin, R. M. (1959): *Toward a Systematic Pragmatics* (Amsterdam: North-Holland).

[258] Martin, Robert (1991): *The Philosopher's Dictionary* (Peterborough, ON: Broadview Press).

[259] Martinich, A. P. & Sosa, David (eds.) (2001): *A Companion to Analytic Philosophy* (Oxford: Blackwell Publishers).

[260] Martin-Löf, Per (1966): "The definition of random sequences", *Information and Control* **9**, 602–619.

[261] Martin-Löf, Per (1970a): *Statistiska modeller*. Seminar notes from the academic year 1969–1970. University of Stockholm, Department of Mathematics.

[262] Martin-Löf, Per (1970b): "On the notion of randomness" in A. Kino *et al.* (eds.): *Intuitionism and Proof Theory*, 73–78 (Amsterdam: North Holland).

[263] Martin-Löf, Per (1971): "Complexity oscillations in infinite binary sequences", *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* **19**, 225–230.

[264] McCall, Storrs (1976): "Objective Time Flow", *Philosophy of Science*, **43** (1976), 337–362. Reprinted in M. Tooley (ed.) (1999).

[265] McDonald, Patrick J. (2002): "Helmholtz's Methodology of Sensory Science, the ZEICHENTHEORIE, and Physical Models of Hearing Mechanism" in M. Heidelberger & F. Stadler (eds.) (2002), 159–183.

[266] McGuinness, Brian (ed.) (1987): *Unified Science — The Vienna Circle Monograph Series originally edited by Otto Neurath, now in an English edition* (Dordrecht: D. Reidel).

[267] McKinsey, J.C.C.; Sugar, A.C. & Suppes, P. (1953): "Axiomatic foundations of classical particle mechanics", *Journal of Rational Mechanics and Analysis* **2**, 253-272.

[268] Mellor, D.H. (1995): *The Facts of Causation* (London: Routledge).

[269] Mill, John Stuart (1924): *Autobiography of John Stuart Mill*, edited by John Jacob Joss, (New York: Columbia University Press).

[270] Mitzman, Arthur (1986): *Sociology and estrangement: three sociologists of Imperial Germany* (New Brunswick, New Jersey: Transaction Publishers).

[271] Monk, Ray (2006): "Getting Inside Heisenberg's Head" in A. L. Macfie (ed.) (2006).

[272] Monk, R. & Palmer, A. (eds.): *Bertrand Russell and the Origins of Analytic Philosophy* (Bristol: Thoemmes).

[273] Moore, G.E. (1993) [1889]: "The Nature of Judgement" in *Selected Writings*, ed. T. Baldwin (London: Routledge), 1–19.

[274] Moore, G.E. (1959): *Philosophical Papers* (London: George Allen and Unwin).

[275] Moore, Gregory H. (1982): *Zermelo's Axiom of Choice – Its Origins, Development and Influence* (New York: Springer-Verlag).

[276] Mormann, Thomas (2000): *Rudolf Carnap* (München: Verlag C.H. Beck).

[277] Mormann, Thomas (2003): "Synthetic Geometry and *Aufbau*" in T. Bonk (ed.) (2003), 45–64.

[278] Mormann, Thomas (2010): "Carnap's Boundless Ocean of Unlimited Possibilities : Between Enlightenment and Romanticism" (preprint).

[279] Morscher, Edgar (2009): "Bernard Bolzano", Stanford Encyclopedia of Philosophy (SEP). Retrieved September 26, 2009 from `http://plato.stanford.edu/entries/bolzano/`.

[280] Morscher, Edgar; Neumaier, Otto & Simons, Peter (eds.) (1998): *Applied Ethics in a Troubled World* (Dordrecht: Kluwer Academic Publishers).

[281] Nagel, Ernst (1961): *The structure of science : problems in logic of scientific explanation* (New York : Hartcourt, Brace World).

[282] Nedo, Michael & Ranchetti, Michele (eds.) (1983): *Wittgenstein: Sein Leben in Bildern und Texten* (Frankfurt am Main: Suhrkamp).

[283] Neurath, Otto; Carnap, Rudolf & Morris, Charles (eds.) (1955): *International Encyclopedia of Unified Science*, Volume I, Nos. 1–5 (Chicago, Illinois: The University of Chicago Press).

[284] Nietzsche, Friedrich (1874): "Vom Nutzen und Nachteil der Historie für das Leben" [1874], *Unzeitgemässe Betrachtungen*, 2nd ed., ed. Peter Pütz (Munich: Goldmann) (1992). Translated by R.J. Hollingdale as "On the Uses and Disadvantages of History for Life" in [Breazeale (1983)].

[285] Niiniluoto, Ilkka & Koskinen, Heikki J. (eds.) (2002): *Wienin piiri* (Helsinki: Gaudeamus).

[286] Niven, William & Jordan, James (eds.) (2003): *Politics and Culture in Twentieth-Century Germany* (Rochester: Camden House).

[287] Noll, Richard (1994): *The Jung Cult* (Princeton: Princeton University Press).

[288] Nozick, Robert (1981): *Philosophical Explanations* (Oxford: Clarendon Press).

[289] O'Neill, Michael J. (2006): "On the Role of Time in Collingwood's Thought" in A. L. Macfie (ed.) (2006).

[290] Parker, Christopher (2000): *The English Idea of History from Coleridge to Collingwood* (Hampshire: Ashgate).

[291] Paty, Michel (1998): *D'Alembert — ou la raison physico-mathématique au siècle des Lumières* (Paris: Les Belles Lettres).

[292] Pearl, Judea (2000): *Causality* (Cambridge: Cambridge University Press).

[293] Pietarinen, Ahti-Veikko (2009): "Logiikka matematiikassa ja filosofiassa Hintikan tapaan" in Manninen & Vilkko (2009), 19–28.

[294] Pippin, Robert (1982): *Kant's Theory of Form. An Essay on the* CRITIQUE OF PURE REASON (New Haven: Yale University Press).

[295] von Plato, Jan (1994): *Creating Modern Probability* (Cambridge: Cambridge University Press).

[296] Pleines, Jürgen-Eckardt (2000): *Bildung im Umbruch* (Hildesheim: Georg Olms Verlag).

[297] Poincaré, Henri (1913): *Dernières Pensées* (Paris: Flammarion).

[298] Popper, Karl (1982): *The Open Universe — From the Postscript to the* LOGIC OF SCIENTIFIC DISCOVERY (Totowa, NJ: Rowman and Littlefield).

[299] Popper, Karl R. (1983): *Realism and the Aim of Science — From the Postscript to the* LOGIC OF SCIENTIFIC DISCOVERY (London: Hutchinson).

[300] Popper, Karl R. & Eccles, John C. (1983) [1977]: *The Self and Its Brain* (London: Routledge & Kegan Paul [Originally published by Springer Verlag, Berlin]).

[301] Rantala, Veikko (1977): *Aspects of Definability*, vol. **29**, Nos. 2–3 of *Acta Philosophica Fennica* (Amsterdam: North-Holland Publishing Company).

[302] Rantala, Veikko & Virtanen, Ari: *Johdatus modaalilogiikkaan* (Helsinki: Gaudeamus).

[303] Reck, Erich H. (ed.) (2002): *From Frege to Wittgenstein* (Oxford: Oxford University Press).

[304] Reck, Erich H. (2004): "From Frege and Russell to Carnap: Logic and Logicism in the 1920s" in S. Awodey & C. Klein (eds.) (2004), 151–180.

[305] Reck, Erich H. (2007): "Carnap and modern logic" in M. Friedman & R. Creath (eds.) (2007), 176–199.

[306] Rédei, Miklós & Stöltzner, Michael (eds.) (2001): *John von Neumann and the Foundations of Quantum Physics* (Dordrecht: Kluwer Academic Publishers).

[307] Redwood, John (1976): *Reason, Ridicule and Religion. The Age of Enlightenment in England 1660–1750* (London: Thames & Hudson Ltd.).

[308] Reichenbach, Hans (1938): *Experience and Prediction* (Chicago: The University of Chicago Press).

[309] Reichenbach, Hans (1946): *Philosophic Foundations of Quantum Mechanics* (Berkeley: University of California Press).

[310] Reichenbach, Hans (1949): *The Theory of Probability*, English translation by Enest H. Hutten & Maria Reichenbach (Berkeley: University of California Press).

[311] Reichenbach, Hans (1951): *The Rise of Scientific Philosophy* (Berkeley: University of California Press).

[312] Reichenbach, Hans (1956): *The Direction of Time*, edited by Maria Reichenbach (Berkeley: University of California Press).

[313] Reichenbach, Hans (1957): *The Philosophy of Space & Time*, translated by Maria Reichenbach & John Freund; with introductory remarks by Rudolf Carnap (New York: Dover Publications, Inc.).

[314] Reichenbach, Hans (1965) [1920]: *The Theory of Relativity and A Priori Knowledge*, translated and edited, with an introduction by Maria Reichenbach (Berkeley: University of California Press).

[315] Reichenbach, Hans (1969): *Axiomatization of the Theory of Relativity*, translated and edited by Maria Reichenbach; foreword by Wesley C. Salmon (Berkeley: University of California Press).

[316] Reisch, George A. (1991): "Did Kuhn kill logical empiricism?", *Philosophy of Science* **58**, 264–277.

[317] Reisch, George A. (2005): *How the Cold War Transformed Philosophy of Science — To the Icy Slopes of Logic* (Cambridge: Cambridge University Press).

[318] Rescher, Nicholas (ed.) (1966): *The Logic of Decision and Action* (Pittsburgh: University of Pittsburgh Press).

[319] Rescher, Nicholas (1973): *The Coherence Theory of Truth* (Oxford at the Clarendon Press).

[320] Rescher, Nicholas (1975): *A Theory of Possibility — A Constructivist and Conceptualistic Account of Possible Individuals and Possible Worlds* (Oxford: Basil Blackwell).

[321] Rescher, Nicholas (1979): *Cognitive Systematization — A systems-theoretic approach to a coherentist theory of knowledge* (Oxford: Basil Blackwell).

[322] Rescher, Nicholas (2001): *Cognitive Pragmatism — The Theory of Knowledge in Pragmatic Perspective* (Pittsburgh: University of Pittsburgh Press).

[323] Richards, Robert J. (2002): *The Romantic Conception of Life — Science and Philosophy in the Age of Goethe* (Chicago: The University of Chicago Press).

[324] Richardson, Alan W. (1998): *Carnap's Construction of the World — The* AUFBAU *and the Emergence of Logical Empiricism* (Cambridge: Cambridge University Press).

[325] Richardson, Alan W. (2007): " 'That Sort of Everyday Image of Logical Positivism': Thomas Kuhn and the Decline of Logical Empiricist Philosophy of Science" in Richardson & Uebel (eds.) (2007), 346–369.

[326] Richardson, Alan & Uebel, Thomas (eds.) (2007): *The Cambridge Companion to Logical Empiricism* (Cambridge: Cambridge University Press).

[327] Rickert, Heinrich (1892): *Der Gegenstand der Erkenntnis*, Fifth edition 1921 (Tübingen: Mohr).

[328] Rickert, Heinrich (1921): *Allgemeinen Grundlegung der Philosophie* (Tübingen: Mohr).

[329] Ricketts, Thomas (1996): "Carnap: From Logical Syntax to Semantics" in R. Giere & A. Richardson (eds.) (1996), 231–250.

[330] Robbins, Dorothy (2001): *Vygotsky's Psychology-Philosophy – A Metaphor for Language Theory and Learning* (New York: Kluwer Academic/Plenum Publishers).

[331] Rorty, Richard; Schneewind, J. B. & Skinner, Quentin (eds.) (1984): *Philosophy in History — Essays on the historiography of philosophy* (Cambridge: Cambridge University Press).

[332] Russell, Bertrand (1992) [1903]: *The Principles of Mathematics*, 2nd edition 1937, (London: Routledge).

[333] Russell, Bertrand (1905): "On Denoting", *Mind* **14**, 479–493. Reprinted in Russell (1956).

[334] Russell, Bertrand (1922) [1914]: *Our Knowledge of External World – as a Field for Scientific Method in Philosophy*, originally published by The Open Court Publishing Company (London: George Allen & Unwin Ltd.)

[335] Russell, Bertrand (1917): *Mysticism and logic*, edited by Robert Charles Marsh (London: Unwin Hyman Ltd.).

[336] Russell, Bertrand (1956) [1918]: "The Philosophy of Logical Atomism" in B. Russell (1956), 177–281.

[337] Russell, Bertrand (1956) [1924]: "Logical Atomism" in B. Russell (1956), 321–343.

[338] Russell, Bertrand (1927): *The Analysis of Matter* (London: Kegan Paul).

[339] Russell, Bertrand (1953): "On the Notion of Cause with Applications to the Free-Will Problem" in H. Feigl M. Brodbeck (eds.) (1953): *Readings in the Philosophy of Science* (New York: Appleton-Century-Crofts).

[340] Russell, Bertrand (1956): *Logic and Knowledge – Essays 1901–1950*, edited by Robert Charles Marsh (London: Unwin Hyman Ltd.)

[341] Russell, Bertrand (1985) [1959]: *My Philosophical Development* (London: Unwin Paperbacks).

[342] Ryckman, Thomas (2005): *The Reign of Relativity — Philosophy in Physics 1915–1925* (Oxford: Oxford University Press).

[343] Ryle, Gilbert (1937): "Taking Sides in Philosophy" in G. Ryle (1971b).

[344] Ryle, Gilbert (1971a): *Collected Papers. Volume I* (London: Hutchison).

[345] Ryle, Gilbert (1971b): *Collected Papers. Volume II* (London: Hutchison).

[346] Sauer, Tilman (2002): "Hopes and Disappointments in Hilbert's Axiomatic 'Foundations of Physics' " in M. Heidelberger & F. Stadler (eds.) (2002), 225–237.

[347] Savitt, Steven F. (ed.) (1995): *Time's Arrow Today — Recent physical and philosophical work on the direction of time* (Cambridge: Cambridge University Press).

[348] Scheuerl, Hans (ed.) (1979): *Klassiker der Pädagogik*, 2 volumes (München: C. H. Beck).

[349] Schilpp, Paul (ed.) (1963): *The Philosophy of Rudolf Carnap* (La Salle, Illinois: Open Court).

[350] Schlick, Moritz (1920): *Raum und Zeit in der gegenwärtigen Physik — Zur Eiführung in das Verständnis der Relativitäts- und Gravitationstheorie* (Berlin: Verlag von Julius Springer).

[351] Schlick, Moritz (1925): *Allgemeine Erkenntnislehre*, zweite Auflage (Berlin: Verlag von Julius Springer).

[352] Schlick, Moritz (1987): *The Problems of Philosophy in Their Interconnection – Winter Semester Lectures, 1933-34*, edited by Henk L. Mulder, A. J. Kox and Rainer Hegselmann; trans. Peter Heath, (Dordrecht: D. Reidel).

[353] Shapere, Dudley (1974): "Scientific Theories and Their Domains" in F. Suppe (ed.) (1977).

[354] Shapere, Dudley (1984): *Reason and the Search for Knowledge* (Dordrecht: D. Reidel).

[355] Smith, George E. (2002): "The Methodology of the *Principia*" in I. Bernard Cohen & George E. Smith (eds.) (2002), 138–173.

[356] Soames, Scott (2003): *Philosophical Analysis in the Twentieth Century*, vol. I (Princeton: Princeton University Press).

[357] Stadler, Friedrich (1997): *Studien zum Wiener Kreis* (Frankfurt-am-Main: Suhrkamp).

[358] Stadler, Friedrich (2007): "The Vienna Circle — Context, Profile, and Development" in A. Richardson & T. Uebel (eds.) (2007), 13–40.

[359] Stadler, Friedrich & Stöltzner, Michael (eds.) (2006): *Time and History — Proceedings of the 28. International Ludwig Wittgenstein Symposium, Kirchberg am Wechsel, Austria 2005* (Frankfurt: Ontos Verlag).

[360] Staiger, Emil (1946) [1991]: *Basic Concepts of Poetics (Grundbegriffe der Poetik)*, translated by Janette C. Hudson and Luanne T. Frank, edited by Marianne Burkhard and Luanne T. Frank, with an introduction by Luanne T. Frank (University Park, Pennsylvania: The Pennsylvania State University Press).

[361] Staiger, Emil (1957): *Goethe * 1749–1786* (Zürich: Atlantis Verlag).

[362] Stanford, Michael (1998): *An Introduction to the Philosophy of History* (Oxford: Blackwell).

[363] Stein, Howard (2002): "Newton's Metaphysics" in I. B. Cohen & G. E. Smith (eds.) (2002), 256–307.

[364] Stein, Howard (2004): "The Enterprise of Understanding and the Enterprise of Knowledge", *Synthese*, **140**, 135–176.

[365] Stenius, Erik (1960): *Wittgenstein's* TRACTATUS (Oxford: Blackwell).

[366] Sterret, Susan G. (2002): "Physical Pictures: Engineering Models circa 1914 and in Wittgenstein's TRACTATUS" in M. Heidelberger & F. Stadler (eds.) (2002), 121–135.

[367] Stroll, A. (2000): *Twentieth-Century Analytic Philosophy* (New York: Columbia University Press).

[368] Suppe, Frederick (ed.) (1977): *The Structure of Scientific Theories* (Urbana, Illinois: University of Illinois Press).

[369] Suppes, Patrick (1969): *Studies in the Methodology of Science* (Dordrecht: D. Reidel).

[370] Suppes, Patrick (1970): *A Probabilistic Theory of Causality* (Amsterdam: North Holland).

[371] Suppes, Patrick (1974): *Probabilistic Metaphysics* (Uppsala: Uppsala Universitet).

[372] Suppes, Patrick (1984): *Probabilistic Metaphysics* (Oxford: Basil Blackwell).

[373] Suppes, Patrick (1987): "Propensity representations of probability", *Erkenntnis*, **26** (1987), 335-358. Reprinted in Suppes, Patrick (1993).

[374] Suppes, Patrick (1993): *Models and Methods in the Philosophy of Science: Selected Essays* (Dordrecht: Kluwer Academic Publishers).

[375] Sweet, William (ed.) (2004): *The Philosophy of History: A Re-examination* (Hampshire: Ashgate).

[376] Szaniawski, Klemens (ed.) (1989): *The Vienna Circle and the Lvov-Warsaw School* (Dordrecht: Kluwer Academic Publishers).

[377] Tait, W. (1997): *Early Analytic Philosophy: Frege, Russell, Wittgenstein* (LaSalle, Illinois: Open Court).

[378] Teilhard de Chardin, Pierre (1959): *The Phenomenon of Man*. English translation by Bernard Wall and Introduction by Julian Huxley (New York: Harper & Row, Publishers).

[379] Tonelli, Giorgio (1994): *Kant's* CRITIQUE OF PURE REASON *within the Tradition of Modern Logic* (Zürich: Georg Olms Verlag Hildesheim).

[380] Tooley, Michael (ed.) (1999): *Analytical Metaphysics — A Collection of Essays* (New York: Garland Publishing, Inc.).

[381] Toulmin, Stephen E. (1953): *The Philosophy of Science* (London: Hutchison University Library).

[382] Toulmin, Stephen E. (1961): *Foresight and Understanding* (London: Hutchison University Library).

[383] Toulmin, Stephen E. (1972): *Human Understanding*. Vol. I. (Princeton, New Jersey: Princeton University Press).

[384] Toulmin, Stephen E. (2001): *Return to Reason* (Cambridge, Massachusetts: Harvard University Press).

[385] Uebel, Thomas E. (ed.) (1991): *Rediscovering the Forgotten Vienna Circle — Austrian Studies on Otto Neurath and the Vienna Circle* (Dordrecht: Kluwer Academic Publishers).

[386] Urmson, J.O. (1956): *Philosophical Analysis: Its Development between the Two World Wars* (Oxford: Oxford University Press).

[387] Vaihinger, Hans (1927): *Die Philosophie des Als Ob — System der Theoretischen, Praktischen und Religiösen Fiktionen der Menschheit* (Leipzig: Felix Meiner Verlag).

[388] Valtonen, Mauri & Karttunen, Hannu (2006): *The Three-Body Problem* (Cambridge: Cambridge University Press).

[389] Vasold, Manfred (1996): "Städtewachstum und Stadterneuerung" in *Das 19. Jahhundert. Ein Lesebuch zur Deutschen Gesichte 1815–1918*, ed. Wolfgang Piereth, 47–51. (Munich: C.H. Beck).

[390] Viljanen, Valtteri (2007): *Spinoza's Dynamics of Being — The Concept of Power and Its Role in Spinoza's Metaphysics*, academic dissertation, (Turku: University of Turku, Department of Philosophy).

[391] Vinen, Richard (2000): *A History in Fragments — Europe in the Twentieth Century* (London: Little, Brown and Company).

[392] Voegelin, Eric (1998): *History of Political Ideas. Volume VI — Revolution and the New Science*. The Collected Works of Eric Voegelin, Volume 24, edited with an introduction by Barry Cooper (Columbia, Missouri: University of Missouri Press).

[393] Waismann, Friedrich (1984): *Wittgenstein und der Wiener Kreis*. Werkausgabe Band 3, Gespräche, aufgezeichnet von Friedrich Waismann (Frankfurt am Main: Suhrkamp).

[394] Wang, Hao (1963): *A Survey of Mathematical Logic* (Peking: Science Press/Amsterdam: North-Holland Publishing).

[395] Wang, Hao (1986): *Beyond Analytic Philosophy* (Cambridge, Massachusetts: The MIT Press).

[396] Wang, Hao (1987): *Reflections on Kurt Gödel* (Cambridge, Massachusetts: The MIT Press).

[397] Wang, Hao (1990): *Computation, Logic, Philosophy — A Collection of Essays* (Peking: Science Press/Dordrecht: Kluwer Academic Publishers).

[398] Weber, Max (1919a): 'Wissenschaft als Beruf' in *Gesammelte Aufsätze zur Wissenschaftslehre*, 4th ed. (1973), 582–613 (Tübingen: Mohr).

[399] Weber, Max (1919b): 'Politik als Beruf' in *Gesammelte Politische Schriften*, edited by J. Winckelmann, 5th ed. (1988), 505-560. (Tübingen: Mohr).

[400] Weyl, Hermann (1949): *Philosophy of Mathematics and Natural Science* (Princeton, New Jersey: Princeton University Press).

[401] Whitehead, Alfred North (1929): *The Function of Reason* (Boston: Beacon Press).

[402] Whitehead, Alfred North (1946): *Science and the Modern World* (Cambridge at the University Press).

[403] Whittaker, E.T. (1937): *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies — with an introduction to the problem of three bodies*. Fourth edition, first edition published in 1904. (Cambridge: Cambridge University Press).

[404] Wiener, Norbert (1956): *I Am a Mathematician — The Later Life of a Prodigy* (New York: Doubleday & Company, Inc.).

[405] Williamson, Timothy (1994): *Vagueness* (London: Routledge).

[406] Wilson, Mark (2006): *Wandering Significance. An Essay on Conceptual Behavior* (Oxford: Clarendon Press).

[407] Wisdom, John (1931): *Interpretation and Analysis in Relation to Bentham's Theory of Definition* (London: Kegan Paul).

[408] Wittgenstein, Ludwig (1953): *Philosophische Untersuchungen = Philosophical investigations*, transl. by G. E. M. Anscombe (Oxford: Blackwell).

[409] Wittgenstein, Ludwig (1958): *Blue and Brown Books*, transl. by G. E. M. Anscombe (Oxford: Blackwell).

[410] Wittgenstein, Ludwig (1979) [1914–1916]: *Notebooks 1914–1916*, second edn. (first edn. (1961)), edited by G.E.M. Anscombe & G.H. von Wright, transl. by G. E. M. Anscombe (Oxford: Blackwell).

[411] Wittgenstein, Ludwig (1980): *Remarks on the Philosophy of Psychology. Volume I*, edited by G.E.M. Anscombe & G.H. von Wright, transl. by G. E. M. Anscombe (Oxford: Blackwell).

[412] Wolpert, David H. (2003): "Collective Intelligence", a working paper, `http://ti.arc.nasa.gov/m/profile/dhw/papers/26.pdf`.

[413] Wolpert, David H. (2008): "An Introduction to Collective Intelligence", a working paper, `http://arxiv.org/pdf/cs.LG/9908014.pdf`.

[414] Woolf, Virginia: *Granite and rainbow : essays* (New York : Harcourt Brace Jovanovich).

[415] von Wright, Georg Henrik (1945): *Looginen empirismi — Eräs nykyisen filosofian pääsuuntia* (Helsinki: Kustannusosakeyhtiö Otava).

[416] von Wright, Georg Henrik (1957): *Logik, filosofi och språk* (Helsingfors: Söderström & Co.).